

TOWNSEND and TOWNSEND and CREW LLP  
Two Embarcadero Center, 8th Floor  
San Francisco, CA 94111-3834  
(415) 576-0200

Atty. Docket No. 023070-068910 **08/785532**

"Express Mail" Label No. EM530419608US

Date of Deposit January 17, 1997

BOX SEQUENCE  
ASST. COMMISSIONER FOR PATENTS  
Washington, D. C. 20231

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Asst. Commissioner for Patents, Washington, D. C. 20231

Sir:

Transmitted herewith for filing is the

- ☐ patent application of  
☐ design patent application of  
☒ continuation-in-part patent application of

By

Sharon Piatt

Inventor(s): 1-00 et al  
Joe W. Gray, Colin Collins, Soo-in Hwang, Tony Godfrey, David Kowbel and Johanna Rommens

For: GENES FROM THE 20q13 AMPLICON AND THEIR USES

☒ This application claims priority from each of the following Application Nos./filing dates:

08/680,395 /7/15/96 ; 08/731,499 /10/16/96 ; \_\_\_\_\_/\_\_\_\_\_.

☐ Please amend this application by adding the following before the first sentence: --This application claims the benefit of U.S. Provisional Application No. 60/\_\_\_\_\_, filed \_\_\_\_\_, the disclosure of which is incorporated by reference.--

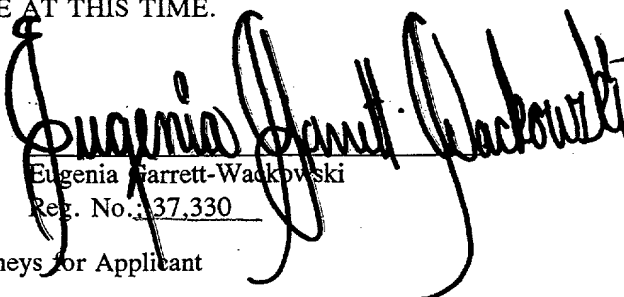
Enclosed are:

- ☒ 6 sheet(s) of ☐ formal ☒ informal drawing(s).  
☐ An assignment of the invention to \_\_\_\_\_  
☐ A ☐ signed ☐ unsigned Declaration & Power of Attorney.  
☒ A ☐ signed ☒ unsigned Declaration.  
☐ A Power of Attorney.  
☐ A verified statement to establish small entity status under 37 CFR 1.9 and 37 CFR 1.27 ☐ is enclosed ☐ was filed in the earliest of the above-identified patent application(s).  
☐ A certified copy of a \_\_\_\_\_ application.  
☐ Information Disclosure Statement under 37 CFR 1.97.  
☐ A petition to extend time to respond in the parent application of this continuation-in-part application.  
☐

In view of the Unsigned Declaration as filed with this application and pursuant to 37 CFR §1.53(d), Applicant requests deferral of the filing fee until submission of the Missing Parts of Application.

DO NOT CHARGE THE FILING FEE AT THIS TIME.

Telephone:  
(415) 576-0200  
APPNOFEE.TRN y  
y11/96

  
Eugenia Farrell-Wackowski  
Reg. No. 37,330  
Attorneys for Applicant

08/785532 A/mf

**TOWNSEND and TOWNSEND and CREW LLP**  
 Two Embarcadero Center, 8th Floor  
 San Francisco, CA 94111-3834  
 (415) 576-0200

Atty. Docket No. 023070-068911

"Express Mail" Label No. EM530419608US

Date of Deposit January 17, 1997

**BOX SEQUENCE**  
**ASST. COMMISSIONER FOR PATENTS**  
 Washington, D. C. 20231

Sir:

Transmitted herewith for filing is the

- ☐ patent application of  
☐ design patent application of  
☒ continuation-in-part patent application of

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Asst. Commissioner for Patents, Washington, D. C. 20231

By

Sharon Piatt

Inventor(s): Joe W. Gray, Colin Collins, Soo-in Hwang, Tony Godfrey, David Kowbel and Johanna Rommens

For: GENES FROM THE 20q13 AMPLICON AND THEIR USES

[x] This application claims priority from each of the following Application Nos./filing dates:

08/680,395 /7/15/96 ; 08/731,499 /10/16/96 ; \_\_\_\_\_/\_\_\_\_\_.  
 08/731,499 /10/16/96 ; \_\_\_\_\_/\_\_\_\_\_.

[ ] Please amend this application by adding the following before the first sentence: --This application claims the benefit of U.S. Provisional Application No. 60/\_\_\_\_\_, filed \_\_\_\_\_, the disclosure of which is incorporated by reference.--

Enclosed are:

- [x] 6 sheet(s) of [ ] formal [x] informal drawing(s).  
 [ ] An assignment of the invention to \_\_\_\_\_  
 [ ] A [ ] signed [ ] unsigned Declaration & Power of Attorney.  
 [X] A [ ] signed [X] unsigned Declaration.  
 [ ] A Power of Attorney.  
 [ ] A verified statement to establish small entity status under 37 CFR 1.9 and 37 CFR 1.27 [ ] is enclosed [ ] was filed in the earliest of the above-identified patent application(s).  
 [ ] A certified copy of a \_\_\_\_\_ application.  
 [ ] Information Disclosure Statement under 37 CFR 1.97.  
 [ ] A petition to extend time to respond in the parent application of this continuation-in-part application.  
 [ ]

In view of the Unsigned Declaration as filed with this application and pursuant to 37 CFR §1.53(d), Applicant requests deferral of the filing fee until submission of the Missing Parts of Application.

DO NOT CHARGE THE FILING FEE AT THIS TIME.

Telephone:  
 (415) 576-0200  
 APPNOFEE.TRN y  
 y11/96

Eugenia Garrett-Wackowski  
 Reg. No.: 37,330

Attorneys for Applicant



## PATENT APPLICATION

### GENES FROM THE 20q13 AMPLICON AND THEIR USES

Inventors:

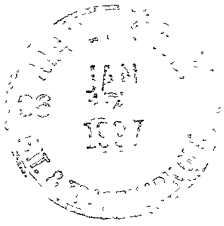
Joe W. Gray  
Colin Collins  
Soo-in Hwang  
Tony Godfrey  
David Kowbel  
Johanna Rommens

Assignee:

Regents of the University of California

Entity:

TOWNSEND and TOWNSEND and CREW  
Two Embarcadero Center, 8th Floor  
San Francisco, CA, 94111-3834



08/785532

1

PATENT

Attorney Docket No. 023070-0689-1-1

**GENES FROM 20q13 AMPLICON AND THEIR USES**

5

**CROSS-REFERENCE TO RELATED APPLICATIONS**

This application is a continuation-in-part of USSN 08/731,499 filed on October 16, 1996 and USSN 08/680,395 filed on July 15, 1996, which is related to copending U.S. Patent Application, USSN 08/546,130, filed October 20, 1995, each of which is incorporated herein by reference for all purposes.

10

**BACKGROUND OF THE INVENTION**

This invention pertains to the field of cytogenetics. More particularly this invention pertains to the identification of genes in a region of amplification at about 20q13 in various cancers. The genes disclosed here can be used as probes specific for the 20q13 amplicon as well as for treatment of various cancers.

15

Chromosome abnormalities are often associated with genetic disorders, degenerative diseases, and cancer. In particular, the deletion or multiplication of copies of whole chromosomes or chromosomal segments, and higher level amplifications of specific regions of the genome are common occurrences in cancer. See, for example Smith, *et al.*, *Breast Cancer Res. Treat.*, 18: Suppl. 1: 5-14 (1991), van de Vijer & Nusse, *Biochim. Biophys. Acta.* 1072: 33-50 (1991), Sato, *et al.*, *Cancer. Res.*, 50: 7184-7189 (1990). In fact, the amplification and deletion of DNA sequences containing proto-oncogenes and tumor-suppressor genes, respectively, are frequently characteristic of tumorigenesis. Dutrillaux, *et al.*, *Cancer Genet. Cytogenet.*, 49: 203-217 (1990).

20

Clearly, the identification of amplified and deleted regions and the cloning of the genes involved is crucial both to the study of tumorigenesis and to the development of cancer diagnostics.

25

The detection of amplified or deleted chromosomal regions has traditionally been done by cytogenetics. Because of the complex packing of DNA into the chromosomes, resolution of cytogenetic techniques has been limited to regions larger than about 10 Mb; approximately the width of a band in Giemsa-stained chromosomes. In complex karyotypes with multiple translocations and other genetic changes, traditional cytogenetic analysis is of little utility because karyotype information is lacking or cannot be interpreted. Teyssier, J.R., *Cancer Genet. Cytogenet.*, 37: 103 (1989). Furthermore,

30

conventional cytogenetic banding analysis is time consuming, labor intensive, and frequently difficult or impossible.

More recently, cloned probes have been used to assess the amount of a given DNA sequence in a chromosome by Southern blotting. This method is effective even if the genome is heavily rearranged so as to eliminate useful karyotype information. However, Southern blotting only gives a rough estimate of the copy number of a DNA sequence, and does not give any information about the localization of that sequence within the chromosome.

Comparative genomic hybridization (CGH) is a more recent approach to identify the presence and localization of amplified/deleted sequences. See Kallioniemi, *et al.*, *Science*, 258: 818 (1992). CGH, like Southern blotting, reveals amplifications and deletions irrespective of genome rearrangement. Additionally, CGH provides a more quantitative estimate of copy number than Southern blotting, and moreover also provides information of the localization of the amplified or deleted sequence in the normal chromosome.

Using CGH, the chromosomal 20q13 region has been identified as a region that is frequently amplified in cancers (*see, e.g.* U.S. Patent No. ). Initial analysis of this region in breast cancer cell lines identified a region approximately 2 Mb on chromosome 20 that is consistently amplified.

## SUMMARY OF THE INVENTION

The present invention relates to the identification of a narrow region (about 600 kb) within a 2 Mb amplicon located at about chromosome 20q13 (more precisely at 20q13.2) that is consistently amplified in primary tumors. In addition, this invention provides cDNA sequences from a number of genes which map to this region. These sequences are useful as probes or as probe targets for monitoring the relative copy number of corresponding sequences from a biological sample such as a tumor cell. Also provided is a contig (a series of clones that contiguously spans this amplicon) which can be used to prepare probes specific for the amplicon. The probes can be used to detect chromosomal abnormalities at 20q13.

Thus, in one embodiment, this invention provides a method of detecting a chromosome abnormality (*e.g.*, an amplification or a deletion) at about position FLpter 0.825 on human chromosome 20 (20q13.2). The method involves contacting a chromosome sample from a patient with a composition consisting essentially of one or

more labeled nucleic acid probes each of which binds selectively to a target polynucleotide sequence at about position FLpter 0.825 on human chromosome 20 under conditions in which the probe forms a stable hybridization complex with the target sequence; and detecting the hybridization complex. The step of detecting the hybridization complex can involve determining the copy number of the target sequence. The probe preferably comprises a nucleic acid that specifically hybridizes under stringent conditions to a nucleic acid selected from the nucleic acids disclosed here. Even more preferably, the probe comprises a subsequence selected from sequences set forth in SEQ. ID. Nos. 1-10 and 12. The probe is preferably labeled, and is more preferably labeled with digoxigenin or biotin. In one embodiment, the hybridization complex is detected in interphase nuclei in the sample. Detection is preferably carried out by detecting a fluorescent label (*e.g.*, FITC, fluorescein, or Texas Red). The method can further involve contacting the sample with a reference probe which binds selectively to a chromosome 20 centromere.

This invention also provides for two new genes, ZABC1 and 1b1, in the 20q13.2 region that are both amplified and overexpressed in a variety of cancers. ZABC1 is a putative zinc finger protein. Zinc finger proteins are found in a variety of transcription factors, and amplification or overexpression of transcription factors typically results in cellular mis-regulation. ZABC1 and 1b1 thus appear to play an important role in the etiology of a number of cancers.

This invention provides for a new human cyclophilin nucleic acid (SEQ ID NO 13). Cyclophilin nucleic acids have been implicated in a variety of cellular processes, including signal transduction.

This invention also provides for proteins encoded by nucleic acid sequences in the 20q13 amplicon (SEQ. ID. Nos: 1-10 and 12-13) and subsequences, more preferably subsequences of at least 10 amino acids, preferably of at least 20 amino acids and most preferably of at least 30 amino acids in length. Particularly preferred subsequences are epitopes specific to the 20q13 proteins, more preferably epitopes specific to the ZABC1 and 1b1 proteins. Such proteins include, but are not limited to isolated polypeptides comprising at least 20 amino acids from a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12-13 or from the polypeptide of SEQ. ID. No. 11 wherein the polypeptide, when presented as an immunogen, elicits the production of an antibody which specifically binds to a polypeptide selected from the group

consisting of a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12-13 or from the polypeptide of SEQ. ID. No. 11, where the polypeptide does not bind to antisera raised against a polypeptide selected from the group consisting of a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12-13 or from the polypeptide of SEQ. ID. No. 11 which has been fully immunosorbed with a polypeptide selected from the group consisting of a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12-13 or from the polypeptide of SEQ. ID. No. 11. In preferred embodiments, the polypeptides of the invention hybridize to antisera raised against a polypeptide encoded by those encoded by SEQ ID NOs. 1-13, where the antisera has been immunosorbed with the most structurally related previously known polypeptide. For example, a polypeptide of the invention binds to antisera raised against a polypeptide encoded by SEQ ID NO. 13, wherein the antisera has been immunosorbed with a rat or mouse cyclophilin polypeptide (Rat cyclophilin nucleic acids are known; *see*, GenBank™ under accession No. M19533; Mouse cyclophilin nucleic acids are known; *see*, GenBank™ under accession No. 50620. cDNAs from the mouse and rat cyclophilin cDNAs are about 85% identical to SEQ ID NO. 13).

In another embodiment, the method can involve detecting a polypeptide (protein) encoded by a nucleic acid (ORF) in the 20q13 amplicon. The method may include any of a number of well known protein detection methods including, but not limited to, the protein assays disclosed herein.

This invention also provides cDNA sequences from genes in the amplicon (SEQ. ID. Nos. 1-10 and 12-13). The nucleic acid sequences can be used in therapeutic applications according to known methods for modulating the expression of the endogenous gene or the activity of the gene product. Examples of therapeutic approaches include antisense inhibition of gene expression, gene therapy, monoclonal antibodies that specifically bind the gene products, and the like. The genes can also be used for recombinant expression of the gene products *in vitro*.

This invention also provides for proteins (*e.g.*, SEQ. ID. No. 11) encoded by the cDNA sequences from genes in the amplicon (*e.g.*, SEQ. ID. Nos. 1-10 and 12-13). Where the amplified nucleic acids include cDNA which are expressed, detection and/or quantification of the protein expression product can be used to identify the presence or absence or quantify the amplification level of the amplicon or of abnormal protein products produced by the amplicon.

The probes disclosed here can be used in kits for the detection of a chromosomal abnormality at about position FLpter 0.825 on human chromosome 20. The kits include a compartment which contains a labeled nucleic acid probe which binds selectively to a target polynucleotide sequence at about FLpter 0.825 on human chromosome 20. The probe preferably includes at least one nucleic acid that specifically hybridizes under stringent conditions to a nucleic acid selected from the nucleic acids disclosed here. Even more preferably, the probes comprise one or more nucleic acids selected from the nucleic acids disclosed here. In a preferred embodiment, the probes are labelled with digoxigenin or biotin. The kit may further include a reference probe specific to a sequence in the centromere of chromosome 20.

### **Definitions**

A "chromosome sample" as used herein refers to a tissue or cell sample prepared for standard *in situ* hybridization methods described below. The sample is prepared such that individual chromosomes remain substantially intact and typically comprises metaphase spreads or interphase nuclei prepared according to standard techniques.

"Nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, would encompass known analogs of natural nucleotides that can function in a similar manner as naturally occurring nucleotides.

An "isolated" polynucleotide is a polynucleotide which is substantially separated from other contaminants that naturally accompany it, e.g., protein, lipids, and other polynucleotide sequences. The term embraces polynucleotide sequences which have been removed or purified from their naturally-occurring environment or clone library, and include recombinant or cloned DNA isolates and chemically synthesized analogues or analogues biologically synthesized by heterologous systems.

"Subsequence" refers to a sequence of nucleic acids that comprise a part of a longer sequence of nucleic acids.

A "probe" or a "nucleic acid probe", as used herein, is defined to be a collection of one or more nucleic acid fragments whose hybridization to a target can be detected. The probe is labeled as described below so that its binding to the target can be detected. The probe is produced from a source of nucleic acids from one or more particular (preselected) portions of the genome, for example one or more clones, an



isolated whole chromosome or chromosome fragment, or a collection of polymerase chain reaction (PCR) amplification products. The probes of the present invention are produced from nucleic acids found in the 20q13 amplicon as described herein. The probe may be processed in some manner, for example, by blocking or removal of repetitive nucleic acids or enrichment with unique nucleic acids. Thus the word "probe" may be used herein to refer not only to the detectable nucleic acids, but to the detectable nucleic acids in the form in which they are applied to the target, for example, with the blocking nucleic acids, *etc.* The blocking nucleic acid may also be referred to separately. What "probe" refers to specifically is clear from the context in which the word is used.

"Hybridizing" refers the binding of two single stranded nucleic acids via complementary base pairing.

"Bind(s) substantially" or "binds specifically" or "binds selectively" or "hybridizes specifically" refer to complementary hybridization between an oligonucleotide and a target sequence and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target polynucleotide sequence. These terms also refer to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA. The term "stringent conditions" refers to conditions under which a probe will hybridize to its target subsequence, but to no other sequences.

Stringent conditions are sequence-dependent and will be different in different circumstances. "Stringent hybridization" and "Stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as CGH, FISH, Southern and northern hybridizations are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* part I chapter 2 "overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York.

Generally, highly stringent hybridization and wash conditions are selected to be about 5° C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at

which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the  $T_m$  for a particular probe.

An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of stringent wash conditions is a .2x SSC wash at 65°C for 15 minutes (*see*, Sambrook, *supra* for a description of SSC buffer). Often, the high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, *e.g.*, about 100 nucleotides or more, is 1x SSC at 45°C for 15 minutes. An example low stringency wash for a duplex of, *e.g.*, more than 100 nucleotides, is 4x SSC at 40°C for 15 minutes. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

One of skill will recognize that the precise sequence of the particular probes described herein can be modified to a certain degree to produce probes that are "substantially identical" to the disclosed probes, but retain the ability to bind substantially to the target sequences. Such modifications are specifically covered by reference to the individual probes herein. The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 90% sequence identity, more preferably at least 95%, compared to a reference sequence using the methods described below using standard parameters.

Two nucleic acid sequences are said to be "identical" if the sequence of nucleotides in the two sequences is the same when aligned for maximum correspondence as described below. The term "complementary to" is used herein to mean that the complementary sequence is identical to all or a portion of a reference polynucleotide sequence. Nucleic acids which do not hybridize to complementary versions of each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two sequences over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as

used herein, refers to a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

5           Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman *Adv. Appl. Math.* 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman *Proc. Natl. Acad. Sci. (U.S.A.)* 85: 2444 (1988), by computerized implementations of these algorithms.

10           "Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by  
15           determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Another indication that nucleotide sequences are substantially identical is if two  
20           molecules hybridize to the same nucleic acid sequence under stringent conditions.

          "Conservatively modified variations" of a particular nucleic acid sequence refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large  
25           number of functionally identical nucleic acids encode any given polypeptide. For instance, the codons CGU, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one  
30           species of "conservatively modified variations." Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by

standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a polypeptide is implicit in each described sequence. Furthermore, one of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

The term "20q13 amplicon protein" is used herein to refer to proteins encoded ORFs in the 20q13 amplicon disclosed herein. Assays that detect 20q13 amplicon proteins are intended to detect the level of endogenous (native) 20q13 amplicon proteins present in subject biological sample. However, exogenous 20q13 amplicon proteins (from a source extrinsic to the biological sample) may be added to various assays to provide a label or to compete with the native 20q13 amplicon protein in binding to an anti-20q13 amplicon protein antibody.. One of skill will appreciate that a 20q13 amplicon protein mimetic may be used in place of exogenous 20q13 protein in this context. A "20q13 protein", as used herein, refers to a molecule that bears one or more 20q13 amplicon protein epitopes such that it is specifically bound by an antibody that specifically binds a native 20q13 amplicon protein.

As used herein, an "antibody" refers to a protein consisting of one or more polypeptides substantially encoded by immunoglobulin genes or fragments of immunoglobulin genes. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon,

which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

The basic immunoglobulin (antibody) structural unit is known to comprise a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain ( $V_L$ ) and variable heavy chain ( $V_H$ ) refer to these light and heavy chains respectively.

Antibodies may exist as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce  $F(ab)'_2$ , a dimer of Fab which itself is a light chain joined to  $V_H-C_H1$  by a disulfide bond. The  $F(ab)'_2$  may be reduced under mild conditions to break the disulfide linkage in the hinge region thereby converting the  $F(ab)'_2$  dimer into an Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region (*see, Fundamental Immunology*, W.E. Paul, ed., Raven Press, N.Y. (1993) for a more detailed description of other antibody fragments). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such Fab' fragments may be synthesized *de novo* either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein also includes antibody fragments either produced by the modification of whole antibodies or synthesized *de novo* using recombinant DNA methodologies.

The phrase "specifically binds to a protein" or "specifically immunoreactive with", when referring to an antibody refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to a protein under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies can be raised to the a 20q13 amplicon protein that bind the 20q13 amplicon protein and not to any other proteins present in a biological sample. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA

immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity.

5

### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1(A) shows disease-free survival of 129 breast cancer patients according to the level of 20q13 amplification. Patients with tumors having high level 20q13 amplification have a shorter disease-free survival ( $p=0.04$  by Mantel-Cox test) compared to those having no or low level amplification.

10

Figure 1(B) Shows the same disease-free survival difference of Figure 4(A) in the sub-group of 79 axillary node-negative patients ( $p=0.0022$  by Mantel-Cox test).

15

Figure 2 shows a comparison of 20q13 amplification detected by FISH in a primary breast carcinoma and its metastasis from a 29-year patient. A low level amplification of 20q13 (20q13 compared to 20p reference probe) was found in the primary tumor. The metastasis, which appeared 8 months after mastectomy, shows a high level amplification of the chromosome 20q13 region. The overall copy number of chromosome 20 (20p reference probe) remained unchanged. Each data point represents gene copy numbers in individual tumor cells analyzed.

20

Figure 3 shows a graphical representation of the molecular cytogenetic mapping and subsequent cloning of the 20q13.2 amplicon. Genetic distance is indicated in centiMorgans (cM). The thick black bar represents the region of highest level amplification in the breast cancer cell line BT474 and covers a region of about 1.5 Mb. P1 and BAC clones are represented as short horizontal lines and YAC clones as heavier horizontal lines. Not all YAC and P1 clones are shown. YACs 957f3, 782c9, 931h12, and 902 are truncated. Sequence tagged sites (STSs) appear as thin vertical lines and open circles indicate that a given YAC has been tested for and is positive for a given STS. Not all STSs have been tested on all YACs. The interval from which more than

25

100 exons have been trapped is represented as a filled box. The 600 kb interval spanning the region of highest amplification level in primary tumors is represented by the filled black box (labeled Sequence). The lower part of the figure shows the levels of amplification in two primary tumors that further narrow the region of highest amplification to within about 600 kb.

30

Figure 4 provides a higher resolution map of the amplicon core as defined in primary tumors.

Figure 5 shows the map location of 15 genes in the amplicon.

Figure 6 shows a sequence alignment between Rat cyclophilin and SEQ

5 ID 13.

### DETAILED DESCRIPTION

This invention provides a number of cDNA sequences which can be used as probes for the detection of chromosomal abnormalities at 20q13. Studies using comparative genomic hybridization (CGH) have shown that a region at chromosome 10 20q13 is increased in copy number frequently in cancers of the breast (~30%), ovary (~15%), bladder (~30%), head and neck (~75%) and colon (~80%). This suggests the presence of one or more genes that contribute to the progression of several solid tumors are located at 20q13.

Gene amplification is one mechanism by which dominantly acting 15 oncogenes are overexpressed, allowing tumors to acquire novel growth characteristics and/or resistance to chemotherapeutic agents. Loci implicated in human breast cancer progression and amplified in 10-25% of primary breast carcinomas include the erbB-2 locus (Lupu *et al.*, *Breast Cancer Res. Treat.*, 27: 83 (1993), Slamon *et al.* *Science*, 235: 177-182 (1987), Heiskanen *et al.* *Biotechniques*, 17: 928 (1994)) at 17q12, cyclin-D 20 (Mahadevan *et al.*, *Science*, 255: 1253-1255 (1993), Gillett *et al.*, *Canc. Res.*, 54: 1812 (1994)) at 11q13 and MYC (Gaffey *et al.*, *Mod. Pathol.*, 6: 654 (1993)) at 8q34.

Pangenomic surveys using comparative genomic hybridization (CGH) recently identified about 20 novel regions of increased copy number in breast cancer (Kallioniemi *et al.*, *Genomics*, 20: 125-128 (1994)). One of these loci, band 20q13, was 25 amplified in 18% of primary tumors and 40% of cell lines (Kallioniemi *et al.*, *Genomics*, 20: 125-128 (1994)). More recently, this same region was found amplified in 15% of ovarian, 80% of bladder and 80% of colorectal tumors. The resolution of CGH is limited to 5-10 Mb. Thus, FISH was performed using locus specific probes to confirm the CGH data and precisely map the region of amplification.

30 The 20q13 region has been analyzed in breast cancer at the molecular level and a region, approximately 600 kb wide, that is consistently amplified was identified, as described herein. Moreover, as shown herein, the importance of this amplification in breast cancer is indicated by the strong association between amplification and decreased

patient survival and increased tumor proliferation (specifically, increased fraction of cells in S-phase).

In particular, as explained in detail in Example 1, high-level 20q13 amplification was associated ( $p=0.0022$ ) with poor disease free survival in node-negative patients, compared to cases with no or low-level amplification (Figure 1). Survival of patients with moderately amplified tumors did not differ significantly from those without amplification. Without being bound to a particular theory, it is suggested that an explanation for this observation may be that low level amplification precedes high level amplification. In this regard, it may be significant that one patient developed a local metastasis with high-level 20q13.2 amplification 8 month after resection of a primary tumor with low level amplification (Figure 3).

The 20q13 amplification was associated with high histologic grade of the tumors. This correlation was seen in both moderately and highly amplified tumors. There was also a correlation ( $p=0.0085$ ) between high level amplification of a region complementary to a particular probe, RMC20C001 (Tanner *et al.*, *Cancer Res.*, 54: 4257-4260 (1994)), and cell proliferation, measured by the fraction of cells in S-phase (Figure 4). This finding is important because it identifies a phenotype that can be scored in functional assays, without knowing the mechanism underlying the increased S-phase fraction. The 20q13 amplification did not correlate with the age of the patient, primary tumor size, axillary nodal or steroid hormone-receptor status.

This work localized the 20q13.2 amplicon to an interval of approximately 2 Mb. Furthermore, it suggests that high-level amplification, found in 7% of the tumors, confers an aggressive phenotype on the tumor, adversely affecting clinical outcome. Low level amplification (22% of primary tumors) was associated with pathological features typical of aggressive tumors (high histologic grade, aneuploidy and cell proliferation) but not patient prognosis.

In addition, it is shown herein that the 20q13 amplicon (more precisely the 20q13.2 amplicon) is one of three separate co-amplified loci on human chromosome 20 that are packaged together throughout the genomes of some primary tumors and breast cancer cell lines. No known oncogenes map in the 20q13.2 amplicon.

#### **Identification of 20q13 Amplicon Probes.**

Initially, a paucity of available molecular cytogenetic probes dictated that FISH probes be generated by the random selection of cosmids from a chromosome 20



specific library, LA20NC01, and map them to chromosome 20 by digital imaging microscopy. Approximately 46 cosmids, spanning the 70 Mb chromosome, were isolated for which fractional length measurements (FLpter) and band assignments were obtained. Twenty six of the cosmids were used to assay copy number in the breast cancer cell line BT474 by interphase FISH analysis. Copy number was determined by counting hybridization signals in interphase nuclei. This analysis revealed that cosmid RMC20C001 (Flpter, 0.824; 20q13.2), described by Stokke *et al.*, *Genomics*, 26: 134-137 (1995), defined the highest-level amplification (~60 copies/cell) in BT474 cells (Tanner *et al.*, *Cancer Res.*, 54: 4257-4260 (1994)).

P1 clones containing genetically mapped sequences were selected from 20q13.2 and used as FISH probes to confirm and further define the region of amplification. Other P1 clones were selected for candidate oncogenes broadly localized to the 20q13.2 region (Flpter, 0.81-0.84). These were selected from the DuPont P1 library (Shepherd, *et al.*, *Proc. Natl. Acad. Sci. USA*, 92: 2629 (1994), available commercially from Genome Systems), by PCR (Saiki *et al.*, *Science*, 230: 1350 (1985)) using primer pairs developed in the 3' untranslated region of each candidate gene. Gene specific P1 clones were obtained for, protein tyrosine phosphatase (PTPN1, Flpter 0.78), melanocortin 3 receptor (MC3R, Flpter 0.81), phosphoenolpyruvate carboxy kinase (PCK1, Flpter 0.85), zinc finger protein 8 (ZNF8, Flpter 0.93), guanine nucleotide-binding protein (GNAS 1, Flpter .873), src-oncogene (SRC, Flpter 0.669), topoisomerase 1 (TOP1, Flpter 0.675), the bcl-2 related gene bcl-x (Flpter 0.526) and the transcription factor E2F-1 (FLpter 0.541). Each clone was mapped by digital imaging microscopy and assigned Flpter values. Five of these genes (SRC, TOPO1, GNAS1, E2F-1 and BC1-x) were excluded as candidate oncogenes in the amplicon because they mapped well outside the critical region at Flpter 0.81-0.84. Three genes (PTPNR1, PCK-1 and MC3R) localized close enough to the critical region to warrant further investigation.

Interphase FISH on 14 breast cancer cell lines and 36 primary tumors using 24 cosmid and 3 gene specific P1 (PTPNR1, PCK-1 and MC3R) probes found high level amplification in 35% (5/14) of breast cancer cell lines and 8% (3/36) of primary tumors with one or more probe. The region with the highest copy number in 4/5 of the cell lines and 3/3 primary tumors was defined by the cosmid RMC20C001. This indicated that PTPNR1, PCK1 and MC3R could also be excluded as candidates for

oncogenes in the amplicon and, moreover, narrowed the critical region from 10 Mb to 1.5-2.0 Mb (*see, Tanner et al., Cancer Res., 54: 4257-4260 (1994).*

Because probe RMC20C001 detected high-level (3 to 10-fold) 20q13.2 amplification in 35% of cell lines and 8% of primary tumors it was used to (1) define the prevalence of amplification in an expanded tumor population, (2) assess the frequency and level of amplification in these tumors, (3) evaluate the association of the 20q13.2 amplicon with pathological and biological features, (4) determine if a relationship exists between 20q13 amplification and clinical outcome and (5) assess 20q13 amplification in metastatic breast tumors.

As detailed in Example 1, fluorescent *in situ* hybridization (FISH) with RMC20C001 was used to assess 20q13.2 amplification in 132 primary and 11 recurrent breast tumors. The absolute copy number (mean number of hybridization signals per cell) and the level of amplification (mean number of signals relative to the p-arm reference probe) were determined. Two types of amplification were found: Low level amplification (1.5-3 fold with FISH signals dispersed throughout the tumor nuclei) and high level amplification (> 3 fold with tightly clustered FISH signals). Low level 20q13.2 amplification was found in 29 of the 132 primary tumors (22%), whereas nine cases (6.8%) showed high level amplification.

RMC20C001 and four flanking P1 probes (MC3R, PCK, RMC20C026, and RMC20C030) were used to study the extent of DNA amplification in highly amplified tumors. Only RMC20C001 was consistently amplified in all tumors. This finding confirmed that the region of common amplification is within a 2 Mb interval flanked by but not including PCK-1 and MC3R.

A physical map was assembled to further localize the minimum common region of amplification and to isolate the postulated oncogene(s). The DuPont P1 library (Shepherd *et al. Proc. Natl. Acad. Sci. USA*, 91: 2629 (1994) was screened for STSs likely to map in band 20q13.2. P1 clones at the loci D20S102, D20S100, D20S120, D20S183, D20S480, D20S211 were isolated, and FISH localized each to 20q13.2. Interphase FISH analysis was then performed in the breast cancer cell line BT474 to assess the amplification level at each locus. The loci D20S100-D20S120-D20S183-D20S480-D20S211 were highly amplified in the BT474 cell line, whereas D20S102 detected only low level amplification. Therefore, 5 STSs,

spanning 5 cM, were localized within the 20q13.2 amplicon and were utilized to screen the CEPH megaYAC library.

CEPH megaYAC library screening and computer searches of public databases revealed D20S120-D20S183-D20S480-D20S211 to be linked on each of three megaYAC clones y820f5, 773h10, and 931h6 (Figure 3). Moreover, screening the CEPH megaYAC library with STSs generated from the ends of cosmids RMC20C001, RMC20C30 and RMC20C028 localized RMC20C001 to each of the same three YAC clones. It was estimated, based on the size of the smallest of these YAC clones, that D20S120-D20S 183-RMC20C001-D20S480-D20S211 map into an interval of less than 1.1 Mb. D20S100 was localized 300 kb distal to D20S120 by interphase FISH and to YAC901b12 by STS mapping. The combined STS data made it possible to construct a 12 member YAC contig which spans roughly 4 Mb encompassing the 1.5 Mb amplicon and containing the loci RMC20C030-PCK1-RMC20C001-MC3R-RMC20C026. Each YAC was mapped by FISH to confirm localization to 20q13.2 and to check for chimerism. Five clonal isolates of each YAC were sized by pulsed field gel electrophoresis (PFGE). None of the YACs are chimeric, however, several are highly unstable.

The YAC contig served as a framework from which to construct a 2 MbP1 contig spanning the 20Q13 amplicon. P1 clones offered numerous advantages over YAC clones including (1) stability, (2) a chimeric frequency of less than 1%, (3) DNA isolation by standard miniprep procedures, (4) they make ideal FISH probes, (5) the ends can be sequenced directly, (6) engineered  $\gamma\delta$  transposons carrying bidirectional primer binding sites can be integrated at any position in the cloned DNA (Strathmann *et al.*, *Proc. Natl. Acad. Sci. USA*, 88: 1247 (1991)) (7) P1 clones are the templates for sequencing the human and Drosophila genomes at the LBNL HGC (Palazzolo *et al.* *DOE Human Genome Program, Contractor-Grantee Workshop IV. Santa Fe, New Mexico*, November 13-17 1994).

About 90 P1 clones were isolated by screening the DuPont P1 library either by PCR or filter hybridization. For PCR based screening, more than 22 novel STSs were created by two methods. In the first method, the ends of P1 clones localized to the amplicon were sequenced, STSs developed, and the P1 library screened for walking clones. In the second approach inter-Alu PCR (Nelson *et al.*, 86: 6686-6690 (1989)) was performed on YACs spanning the amplicon and the products cloned and

sequenced for STS creation. In the filter based hybridization scheme P1 clones were obtained by performing inter-Alu PCR on YACs spanning the amplicon, radio-labeling the products and hybridizing these against filters containing a gridded array of the P1 library. Finally, to close gaps a human genomic bacterial artificial chromosome (BAC) library (Shizuya *et al. Proc. Natl. Acad. Sci. USA*, 89: 8794 (1992), commercially available from Research Genetics, Huntsville, Alabama, USA) was screened by PCR. These methods combined to produce more than 100 P1 and BAC clones were localized to 20q13.2 by FISH. STS content mapping, fingerprinting, and free-chromatin fish (Heiskanen *et al., BioTechniques*, 17: 928 (1994)) were used to construct the 2 Mb contig shown in Figure 3.

#### **Fine Mapping the 20q13.2 Amplicon in BT474**

Clones from the 2 Mb P1 contig were used with FISH to map the level of amplification at 20q13.2 in the breast cancer cell line BT474. 35 P1 probes distributed at regular intervals along the contig were used. The resulting data indicated that the region of *highest copy number increase* in BT474 occurs between D20S100 and D20S211, an interval of approximately 1.5 Mb. P1 FISH probes, in this interval, detect an average of 50 signals per interphase nuclei in BT474, while no, or only low level amplification, was detected with the P1 clones outside this region. Thus, both the proximal and distal boundaries of the amplicon were cloned.

#### **Fine Mapping the 20q13.2 Amplicon in Primary Tumors.**

Fine mapping the amplicon in primary tumors revealed the *minimum common region of high amplification* that is of pathobiological significance. This process is analogous to screening for informative meiosis in the narrowing of genetic intervals encoding heritable disease genes. Analysis of 132 primary tumors revealed thirty-eight primary tumors that are amplified at the RMC20C001 locus. Nine of these tumors have high level amplification at the RMC20C001 locus and were further analyzed by interphase FISH with 8 P1s that span the  $\approx 2$  Mb contig. The minimum common region of amplification was mapped to a  $\approx 600$  kb interval flanked by P1 clones #3 and #12 with the highest level of amplification detected by P1 clone #38 corresponding to RMC20C001 (Figure 4).

The P1 and BAC clones spanning the 600 kb interval of the 20q13 amplicon are listed in Table 1 which provides a cross-reference to the DuPont P1 library described by Shepherd, *et al., Proc. Natl. Acad. Sci. USA*, 92: 2629 (1994). These P1

Table 1. Cross reference to probes of the DuPont P1 library (Shepherd, *et al.*, *Proc. Natl. Acad. Sci. USA*, 92:2629 (1994) which is commercially available from Genomic Systems, St. Louis, Missouri, USA). PCR primers are illustrated for amplification of Sequence tag sites for each clone. In addition, PCR conditions (Mg concentration and annealing temperature), as well as PCR product size, is provided. Size: PCR product size; mM MgCl: Mg concentration; Ann.: Annealing temperature; P1: P1 probe ID number; PC: DuPont Library Plate Coordinates; SCA: DuPont Library Single Clone address; SEQ-forward and SEQ-backward: forward and backward PCR primers, respectively; SEQ ID NO:-forward and SEQ ID NO:-backward: Sequence Listing SEQ ID NO: for forward and backward primers, respectively.

PRIMER NAME	SIZE (bp)	mM MgCl	Ann.	P1	PC	SCA	SEQ-forward	SEQ-backward	SEQ ID NO:-forward	SEQ ID NO:-backward
352.32	136	1.5	52	20	103-e5	1228e	TTGGCAATTGGTATCAGGTAGCTG	TTGGAGCAGAGAGGGGATTGTGTG	9	10
388.13 F1/B1	201	1.5	52	17	69g6	821	AATCCCTCAAAACCTGCTGCTAC	TGGAGCCTGAACCTCTGCAATC	11	12
D20S183	270	3	48	30		1167f	CCGGATACCCGACATTG	TGCACATAAAACAGCCAGC	13	14
D20S211 1/2	135	1.5	52	29	24h1	276h	TTGGAATCAATGGAGCAAAA	AGCTTTACCCCAATGTGGTCC	15	16
D20S480	300	3	55	68	100d12	1199d2	GTGGTGAACACCAATAAATGG	AAGCAATAAACCAATAAATCG	17	18
				41	86c1	1020c				
				42	103d9	1232d				
				67	91b2	1081b9				
9X-SP6 hmF/hmB	165	1.5	55	7	31d11	370d	CAAGATCTGACCCGTCATC	GACTTCTTCAGGAAAGATCATG	19	20
11S-17 F2/B4	146	3	58	11	41b1	480b	GCCATGTACCACCTGAAAAATC	TCAGAACACCCGTGCAGAAATTAAG	21	22
12T-17 F2/B1	153	3	58	12	42c2	493c	CCTAAAACTTGGTGTCTTAATCTA	GTCTCAAGGCGAGATGTGG	23	24
28T F1/B1	219	1.5	52	74		888f2	TTGTGTATGTTGAGCCATC	CTTCCAATCTCATTTCTATGAGG	25	26
				2	12c6	137c				
				25	118c11	1413c				
				26	118c11	1413c				
28S F1/B3	214	1.5	55	28	118g11	1413g	GCTTGTAAAGTGTCACTAGGG	CACTCTGGTAAATGACCTTTGTC	27	28
				25	118c11	1413c				
				27	118g11	1413g				
				10	36f10	429f				
69S	100	3	55	69		412b5	CCTACACCATTCCAACITTTGG	GCCAGATGTATGTTGTACGGAAC	29	30
HSCOFH032 F/B	129	1.5	55	5	23c1	264c				
				3	12-e11	142e	TTCTAAACCTGTCCACTTCTTG	CTGCTGTGGTGGAGAATGG	31	32
60A1	191	1.5	58	36	77a10	921a	TGTCCTCCTTGTCCCTCATCTCTAC	AATGCCCTCCACTCACAGGAATG	33	34
820F5A1T F1/B1	175	1.5	48	15	53c7	630c	CTCTTCAGTGTCTTCTCTATTGA	GGGAGGAGGTGTAGGCAAC	35	36
				16	58b9	692b				

Table 2. Cross reference to probes of the BAC library. Clone # refers to the clone number provided, e.g., in Figure 4, while the plate coordinates are the plate coordinates in the Research Genetics BAC library. Size: PCR product Size; mM MgCl: Mg concentration; Ann.: Annealing temperature; BAC #: BAC probe ID number; SEQ-forward and SEQ-backward: forward and backward PCR primers, respectively; SEQ ID NO:-forward and SEQ ID NO:-backward: Sequence Listing SEQ ID NO: for forward and backward primers, respectively.

PRIMER NAME	SIZE (bp)	mM MgCl	Ann.	BAC #	BAC Plate Coordinates	SEQ-forward	SEQ-backward	SEQ ID NO:-forward	SEQ ID NO:-backward
18T F1/B1	156	3	62	99	L11 plate 146	AGCAAAAGCAAAGGTGGCACAC	TGACATGGGAGAAGACACACTTCC	37	38
9S F1/B1	214	1.5	55	97	E8 plate 183	AGGTTTACCAATGTGTTTGG	TCTACATCCCATCTCTCTCG	39	40
D20S480	300	3	55	95	H15 plate 140	GTGGTGAACACCAATAAATGG	AAGCAAAATAAAACCAATAAACTCG	17	18
D20S211 1/2	135	1.5	52	103	A15 plate 188	TTGGAATCAATGGAGCAAAA	AGCTTTACCCCAATGTGTCTCC	15	16
				102	A1 plate 46			*	*
11S-17 F2/B4	146	3	58	100	E4 plate 43	GCCATGTACCCACCTGAAAAATC	TCAGAACACCCCGTGCAGAAATTAAG	21	22
				101	J5 plate 118			*	*

and BAC probes are available commercially from Genetic Systems, and Research Genetics, respectively).

**cDNA sequences from the 20q13 amplicon.**

Exon trapping (*see, e.g.,* Duyk *et al.*, *Proc. Natl. Acad. Sci. USA*, 87: 8995-8999 (1990) and Church *et al.*, *Nature Genetics*, 6: 98-105 (1994)) was performed on the P1 and BAC clones spanning the  $\approx 600$  kb minimum common region of amplification and has isolated more than 200 exons.

Analysis of the exons DNA sequence revealed a number of sequence similarities (85% to 96%) to partial CDNA sequences in the expressed sequence data base (dbest) and to a *S. cerevisiae* chromosome XIV open reading frame. Each P1 clone subjected to exon trapping has produced multiple exons consistent with at least a medium density of genes. Over 200 exons have been trapped and analyzed as well as 200 clones isolated by direct selection from a BT474 cDNA library. In addition a 0.6 Mb genomic interval spanning the minimal amplicon described below is being sequenced. Exon prediction and gene modeling are carried out with XGAIL, SORFIND, and BLAST programs. Gene fragments identified by these approaches have been analyzed by RT-PCR, Northern and Southern blots. Fifteen unique genes were identified in this way (*see*, Table 3 and Figure 5).

In addition two other genes ZABC1 (SEQ. ID. 9 and 10) and 1b1 (SEQ ID No. 12) were also were shown to be overexpressed in a variety of different cancer cells.

Sequence information from various cDNA clones are provided below. They are as follows:

3bf4 (SEQ. ID. No. 1) - 3kb transcript with sequence identity to a tyrosine kinase gene, termed A6, disclosed in Beeler *et al.* *Mol. Cell. Biol.* 14:982-988 (1994) and WO 95/19439. These references, however, do not disclose that the gene is located in the chromosome 20 amplicon.

1b11 (SEQ. ID. No. 2) - an approximately 3.5 kb transcript whose expression shows high correlation with the copy number of the amplicon. The sequence shows no homology with sequences in the databases searched.

cc49 (SEQ. ID. No. 3) - a 6-7 kb transcript which shows homology to C2H2 zinc finger genes.

cc43 (SEQ. ID. No. 4) - an approximately 4 kb transcript which is expressed in normal tissues, but whose expression in the breast cancer cell line has not been detected.

41.1 (SEQ. ID. No. 5) - shows homology to the homeobox T shirt gene in *Drosophila*.

GCAP (SEQ. ID. No. 6) - encodes a guanine cyclase activating protein which is involved in the biosynthesis of cyclic AMP. As explained in detail below, sequences from this gene can also be used for treatment of retinal degeneration.

1b4 (SEQ. ID. No. 7) - a serine threonine kinase.

20sa7 (SEQ. ID. No. 8) - a homolog of the rat gene, BEM-1.

In addition, the entire nucleotide sequence is provided for ZABC-1.

ZABC-1 stands for zinc finger amplified in breast cancer. This gene maps to the core of the 20q13.2 amplicon and is overexpressed in primary tumors and breast cancer cell lines having 20q13.2 amplification. The genomic sequence (SEQ. ID. No. 9) includes roughly 2kb of the promoter region. SEQ ID. No. 10 provides the cDNA sequence derived open reading frame and SEQ ID. No. 11 provides the predicted protein sequence. Zinc finger containing genes are often transcription factors that function to modulate the expression of down stream genes. Several known oncogenes are in fact zinc finger containing genes.

This invention also provides the full length cDNA sequence for a cDNA designated 1b1 (SEQ. ID. No. 12) which is also overexpressed in numerous breast cancer cell lines and some primary tumors.

SEQ ID NO: 13 provides sequence from a genomic clone which is similar to known rat and mouse cyclophilin cDNAs. Rat Cyclophilin nucleic acids (*e.g.*, cDNAs) are known; *see*, GenBank™ under accession No. M19533; Mouse Cyclophilin nucleic acids (*e.g.*, cDNAs) are known; *see*, GenBank™ under accession No. 50620. Accordingly, SEQ ID NO: 13 is a putative human cyclophilin gene. The sequence is also associated with amplified sequences from 20q13, and can be used as a probe or probe hybridization target to detect DNA amplification, or RNA overexpression of the corresponding gene.



**Table 3.** Gene fragments identified by exon trapping and analyzed by RT-PCR, Northern and Southern blots.

Map #	Gene ID	Transcript Size (kb)	EST Identity (>95%)	Cloned Sequence (kb)	Protein Homologies	Map Location
1	20sa7	-	Yes	3	PTP	3, 18, 99
2	1b11	3.5	No	1.5	novel	18, 3, 99, 69
3	200.1	-	-	-	-	-
4	132.1	-	-	-	novel	132
5	132.2	-	-	-	-	132
6	3bf4	3	Yes	3	PTK	ambiguous
7	7.1	-	-	-	-	7, 11, 97
8	7.2	2.4	-	-	-	7, 11, 97
9	cc49	7, 4	Yes	3.6	Kruppel	97, 103
10	cc43	1.4	Yes	1.8	hypothetical yeast protein	97, 7, 11
11	et1807	2.5	Yes	0.7	novel	97, 9
12	et2106	None	Yes	-	-	95, 39, 38
13	41.1	detected	Yes	3	homeotic	95, 41, 42
14	67.1	7, 8, 11	Yes	2Kb	gene	67
15	67.2	-	Yes	-	cGMP reg. protein	67
		0				

#### 20q13 Amplicon Proteins

As indicated above, this invention also provides for proteins encoded by nucleic acid sequences in the 20q13 amplicon (*e.g.*, SEQ. ID. Nos: 1-10 and 12-13) and subsequences more preferably subsequences of at least 10 amino acids, preferably of at least 20 amino acids, and most preferably of at least 30 amino acids in length.

Particularly preferred subsequences are epitopes specific to the 20q13 proteins more preferably epitopes specific to the ZABC1 and 1b1 proteins. Such proteins include, but

are not limited to isolated polypeptides comprising at least 10 contiguous amino acids from a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12-13 or from the polypeptide of SEQ. ID. No. 11 wherein the polypeptide, when presented as an immunogen, elicits the production of an antibody which specifically binds to a polypeptide selected from the group consisting of a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12-13 or from the polypeptide of SEQ. ID. No. 11 and the polypeptide does not bind to antisera raised against a polypeptide selected from the group consisting of a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12 or from the polypeptide of SEQ. ID. No. 11 which has been fully immunosorbed with a polypeptide selected from the group consisting of a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12-13 or from the polypeptide of SEQ. ID. No. 11.

A protein that specifically binds to or that is specifically immunoreactive with an antibody generated against a defined immunogen, such as an immunogen consisting of the amino acid sequence of SEQ ID NO 11 is determined in an immunoassay. The immunoassay uses a polyclonal antiserum which was raised to the protein of SEQ ID NO 11 (the immunogenic polypeptide). This antiserum is selected to have low crossreactivity against other similar known polypeptides and any such crossreactivity is removed by immunoabsorbption prior to use in the immunoassay (*e.g.*, by immunosorbption of the antisera with the related polypeptide).

In order to produce antisera for use in an immunoassay, the polypeptide *e.g.*, the polypeptide of SEQ ID NO 11 is isolated as described herein. For example, recombinant protein can be produced in a mammalian or other eukaryotic cell line. An inbred strain of mice is immunized with the protein of SEQ ID NO 11 using a standard adjuvant, such as Freund's adjuvant, and a standard mouse immunization protocol (see Harlow and Lane, *supra*). Alternatively, a synthetic polypeptide derived from the sequences disclosed herein and conjugated to a carrier protein is used as an immunogen. Polyclonal sera are collected and titered against the immunogenic polypeptide in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Polyclonal antisera with a titer of  $10^4$  or greater are selected and tested for their cross reactivity against known polypeptides using a competitive binding immunoassay such as the one described in Harlow and Lane, *supra*, at pages 570-573. Preferably more than one known polypeptide is used in this determination in conjunction with the immunogenic polypeptide.

The known polypeptides can be produced as recombinant proteins and isolated using standard molecular biology and protein chemistry techniques as described herein.

Immunoassays in the competitive binding format are used for crossreactivity determinations. For example, the immunogenic polypeptide is immobilized to a solid support. Proteins added to the assay compete with the binding of the antisera to the immobilized antigen. The ability of the a proteins to compete with the binding of the antisera to the immobilized protein is compared to the immunogenic polypeptide. The percent crossreactivity for the protein is calculated, using standard calculations. Those antisera with less than 10% crossreactivity to known polypeptides are selected and pooled. The cross-reacting antibodies are then removed from the pooled antisera by immunoabsorbtion with known polypeptide.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described herein to compare a "target" polypeptide to the immunogenic polypeptide. To make this comparison, the two polypeptides are each assayed at a wide range of concentrations and the amount of each polypeptide required to inhibit 50% of the binding of the antisera to the immobilized protein is determined using standard techniques. If the amount of the target polypeptide required is less than twice the amount of the immunogenic polypeptide that is required, then the target polypeptide is said to specifically bind to an antibody generated to the immunogenic protein. As a final determination of specificity, the pooled antisera is fully immunosorbed with the immunogenic polypeptide until no binding to the polypeptide used in the immunosorbtion is detectable. The fully immunosorbed antisera is then tested for reactivity with the test polypeptide. If no reactivity is observed, then the test polypeptide is specifically bound by the antisera elicited by the immunogenic protein.

Similarly, in a reciprocal experiment, the pooled antisera is immusorbed with the test polypeptide. If the antisera which remains after the immusorbtion does not bind to the immunogenic polypeptide (*i.e.*, the polypeptide of SEQ ID NO: 11 used to elicit the antisera) then the test polypeptide is specifically bound by the antisera elicited by the immunogenic peptide.

#### **Detection of 20q13 Abnormalities.**

One of skill in the art will appreciate that the clones and sequence information provided herein can be used to detect amplifications, or other chromosomal

abnormalities, at 20q13 in a biological sample. Generally the methods involve hybridization of probes that specifically bind one or more nucleic acid sequences of the target amplicon with nucleic acids present in a biological sample or derived from a biological sample.

As used herein, a biological sample is a sample of biological tissue or fluid containing cells desired to be screened for chromosomal abnormalities (*e.g.* amplifications of deletions). In a preferred embodiment, the biological sample is a cell or tissue suspected of being cancerous (transformed). Methods of isolating biological samples are well known to those of skill in the art and include, but are not limited to, aspirations, tissue sections, needle biopsies, and the like. Frequently the sample will be a "clinical sample" which is a sample derived from a patient. It will be recognized that the term "sample" also includes supernatant (containing cells) or the cells themselves from cell cultures, cells from tissue culture and other media in which it may be desirable to detect chromosomal abnormalities.

In a preferred embodiment, a biological sample is prepared by depositing cells, either as single cell suspensions or as tissue preparation, on solid supports such as glass slides and fixed by choosing a fixative which provides the best spatial resolution of the cells and the optimal hybridization efficiency.

### **Making Probes**

Any of the P1 probes listed in Table 1, the BAC probes listed in Table 2, or the cDNAs disclosed here are suitable for use in detecting the 20q13 amplicon.

Methods of preparing probes are well known to those of skill in the art (*see, e.g.* Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual* (2nd ed.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989) or *Current Protocols in Molecular Biology*, F.

Ausubel *et al.*, ed. Greene Publishing and Wiley-Interscience, New York (1987))

Given the strategy for making the nucleic acids of the present invention, one of skill can construct a variety of vectors and nucleic acid clones containing functionally equivalent nucleic acids. Cloning methodologies to accomplish these ends, and sequencing methods to verify the sequence of nucleic acids are well known in the art. Examples of appropriate cloning and sequencing techniques, and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology* volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.* (1989) *Molecular*

*Cloning - A Laboratory Manual* (2nd ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY, (Sambrook); and *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1994 Supplement) (Ausubel).

5 Product information from manufacturers of biological reagents and experimental equipment also provide information useful in known biological methods. Such manufacturers include the SIGMA chemical company (Saint Louis, MO), R&D systems (Minneapolis, MN), Pharmacia LKB Biotechnology (Piscataway, NJ), CLONTECH Laboratories, Inc. (Palo Alto, CA), Chem Genes Corp., Aldrich Chemical Company  
10 (Milwaukee, WI), Glen Research, Inc., GIBCO BRL Life Technologies, Inc. (Gaithersburg, MD), Fluka Chemica-Biochemika Analytika (Fluka Chemie AG, Buchs, Switzerland), Invitrogen, San Diego, CA, and Applied Biosystems (Foster City, CA), as well as many other commercial sources known to one of skill.

The nucleic acids provided by this invention, whether RNA, cDNA,  
15 genomic DNA, or a hybrid of the various combinations, are isolated from biological sources or synthesized *in vitro*. The nucleic acids and vectors of the invention are present in transformed or transfected whole cells, in transformed or transfected cell lysates, or in a partially purified or substantially pure form.

*In vitro* amplification techniques suitable for amplifying sequences to  
20 provide a nucleic acid, or for subsequent analysis, sequencing or subcloning are known. Examples of techniques sufficient to direct persons of skill through such *in vitro* amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Q $\beta$ -replicase amplification and other RNA polymerase mediated techniques (*e.g.*, NASBA) are found in Berger, Sambrook, and Ausubel, as well as  
25 Mullis *et al.*, (1987) U.S. Patent No. 4,683,202; *PCR Protocols A Guide to Methods and Applications* (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) *C&EN* 36-47; *The Journal Of NIH Research* (1991) 3, 81-94; (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86, 1173; Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87, 1874; Lomell *et al.* (1989) *J. Clin. Chem* 35,  
30 1826; Landegren *et al.*, (1988) *Science* 241, 1077-1080; Van Brunt (1990) *Biotechnology* 8, 291-294; Wu and Wallace, (1989) *Gene* 4, 560; Barringer *et al.* (1990) *Gene* 89, 117, and Sooknanan and Malek (1995) *Biotechnology* 13: 563-564. Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace *et al.*, U.S. Pat. No.

5,426,039. Improved methods of amplifying large nucleic acids are summarized in Cheng *et al.* (1994) *Nature* 369: 684-685 and the references therein.

Nucleic Acids (*e.g.*, oligonucleotides) for *in vitro* amplification methods or for use as gene probes, for example, are typically chemically synthesized according to the solid phase phosphoramidite triester method described by Beaucage and Caruthers (1981), *Tetrahedron Letts.*, 22(20):1859-1862, *e.g.*, using an automated synthesizer, as described in Needham-VanDevanter *et al.* (1984) *Nucleic Acids Res.*, 12:6159-6168. Purification of oligonucleotides, where necessary, is typically performed by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson and Regnier (1983) *J. Chrom.* 255:137-149. The sequence of the synthetic oligonucleotides can be verified using the chemical degradation method of Maxam and Gilbert (1980) in Grossman and Moldave (eds.) Academic Press, New York, *Methods in Enzymology* 65:499-560.

The probes are most easily prepared by combining and labeling one or more of the constructs listed in Tables 1 and 2. Prior to use, the constructs are fragmented to provide smaller nucleic acid fragments that easily penetrate the cell and hybridize to the target nucleic acid. Fragmentation can be by any of a number of methods well known to those of skill in the art. Preferred methods include treatment with a restriction enzyme to selectively cleave the molecules, or alternatively to briefly heat the nucleic acids in the presence of  $Mg^{2+}$ . Probes are preferably fragmented to an average fragment length ranging from about 50 bp to about 2000 bp, more preferably from about 100 bp to about 1000 bp and most preferably from about 150 bp to about 500 bp.

Alternatively, probes can be produced by amplifying (*e.g.* via PCR) selected subsequences from the 20q13 amplicon disclosed herein. The sequences provided herein permit one of skill to select primers that amplify sequences from one or more exons located within the 20q13 amplicon.

Particularly preferred probes include nucleic acids from probes 38, 40, and 79, which corresponds to RMC20C001. In addition, the cDNAs are particularly useful for identifying cells that have increased expression of the corresponding genes, using for instance, Northern blot analysis.

One of skill will appreciate that using the sequence information and clones provided herein, one of skill in the art can isolate the same or similar probes from other human genomic libraries using routine methods (*e.g.* Southern or Northern Blots).

Similarly, the polypeptides of the invention can be synthetically prepared in a wide variety of well-known ways. For instance, polypeptides of relatively short length can be synthesized in solution or on a solid support in accordance with conventional techniques. *See, e.g.,* Merrifield (1963) *J. Am. Chem. Soc.* 85:2149-2154. Various automatic synthesizers are commercially available and can be used in accordance with known protocols. *See, e.g.,* Stewart and Young (1984) *Solid Phase Peptide Synthesis*, 2d. ed., Pierce Chemical Co. As described in more detail herein, the polypeptide of the invention are most preferably made using recombinant techniques, *e.g.,* by expressing the polypeptides in host cells and purifying the expressed proteins.

In a preferred embodiment, the polypeptides, or subsequences thereof, are synthesized using recombinant DNA methodology. Generally this involves creating a DNA sequence that encodes the protein, through recombinant, synthetic, or *in vitro* amplification techniques, placing the DNA in an expression cassette under the control of a particular promoter, expressing the protein in a host cell, isolating the expressed protein and, if required, renaturing the protein.

#### **Labeling Probes**

Methods of labeling nucleic acids are well known to those of skill in the art. Preferred labels are those that are suitable for use in *in situ* hybridization. The nucleic acid probes may be detectably labeled prior to the hybridization reaction. Alternatively, a detectable label which binds to the hybridization product may be used. Such detectable labels include any material having a detectable physical or chemical property and have been well-developed in the field of immunoassays.

As used herein, a "label" is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. Useful labels in the present invention include radioactive labels (*e.g.*  $^{32}\text{P}$ ,  $^{125}\text{I}$ ,  $^{14}\text{C}$ ,  $^3\text{H}$ , and  $^{35}\text{S}$ ), fluorescent dyes (*e.g.* fluorescein, rhodamine, Texas Red, *etc.*), electron-dense reagents (*e.g.* gold), enzymes (as commonly used in an ELISA), colorimetric labels (*e.g.* colloidal gold), magnetic labels (*e.g.* Dynabeads<sup>TM</sup>), and the like. Examples of labels which are not directly detected but are detected through the use of directly detectable label include

biotin and dioxigenin as well as haptens and proteins for which labeled antisera or monoclonal antibodies are available.

The particular label used is not critical to the present invention, so long as it does not interfere with the *in situ* hybridization of the stain. However, stains directly  
5 labeled with fluorescent labels (*e.g.* fluorescein-12-dUTP, Texas Red-5-dUTP, *etc.*) are preferred for chromosome hybridization.

A direct labeled probe, as used herein, is a probe to which a detectable label is attached. Because the direct label is already attached to the probe, no subsequent steps are required to associate the probe with the detectable label. In contrast, an  
10 indirect labeled probe is one which bears a moiety to which a detectable label is subsequently bound, typically after the probe is hybridized with the target nucleic acid.

In addition the label must be detectible in as low copy number as possible thereby maximizing the sensitivity of the assay and yet be detectible above any background signal. Finally, a label must be chosen that provides a highly localized  
15 signal thereby providing a high degree of spatial resolution when physically mapping the stain against the chromosome. Particularly preferred fluorescent labels include fluorescein-12-dUTP and Texas Red-5-dUTP.

The labels may be coupled to the probes in a variety of means known to those of skill in the art. In a preferred embodiment the nucleic acid probes will be  
20 labeled using nick translation or random primer extension (Rigby, *et al. J. Mol. Biol.*, 113: 237 (1977) or Sambrook, *et al., Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1985)).

One of skill in the art will appreciate that the probes of this invention need not be absolutely specific for the targeted 20q13 region of the genome. Rather, the  
25 probes are intended to produce "staining contrast". "Contrast" is quantified by the ratio of the probe intensity of the target region of the genome to that of the other portions of the genome. For example, a DNA library produced by cloning a particular chromosome (*e.g.* chromosome 7) can be used as a stain capable of staining the entire chromosome. The library contains both sequences found only on that chromosome, and sequences  
30 shared with other chromosomes. Roughly half the chromosomal DNA falls into each class. If hybridization of the whole library were capable of saturating all of the binding sites on the target chromosome, the target chromosome would be twice as bright (contrast ratio of 2) as the other chromosomes since it would contain signal from the both



the specific and the shared sequences in the stain, whereas the other chromosomes would only be stained by the shared sequences. Thus, only a modest decrease in hybridization of the shared sequences in the stain would substantially enhance the contrast. Thus contaminating sequences which only hybridize to non-targeted sequences, for example, impurities in a library, can be tolerated in the stain to the extent that the sequences do not reduce the staining contrast below useful levels.

### **Detecting the 20q13 Amplicon.**

As explained above, detection of amplification in the 20q13 amplicon is indicative of the presence and/or prognosis of a large number of cancers. These include, but are not limited to breast, ovary, bladder, head and neck, and colon.

In a preferred embodiment, a 20q13 amplification is detected through the hybridization of a probe of this invention to a target nucleic acid (*e.g.* a chromosomal sample) in which it is desired to screen for the amplification. Suitable hybridization formats are well known to those of skill in the art and include, but are not limited to, variations of Southern Blots, *in situ* hybridization and quantitative amplification methods such as quantitative PCR (*see, e.g.* Sambrook, *supra.*, Kallioniemi *et al.*, *Proc. Natl Acad Sci USA*, 89: 5321-5325 (1992), and *PCR Protocols, A Guide to Methods and Applications*, Innis *et al.*, Academic Press, Inc. N.Y., (1990)).

### **In situ Hybridization.**

In a preferred embodiment, the 20q13 amplicon is identified using *in situ* hybridization. Generally, *in situ* hybridization comprises the following major steps: (1) fixation of tissue or biological structure to be analyzed; (2) prehybridization treatment of the biological structure to increase accessibility of target DNA, and to reduce nonspecific binding; (3) hybridization of the mixture of nucleic acids to the nucleic acid in the biological structure or tissue; (4) posthybridization washes to remove nucleic acid fragments not bound in the hybridization and (5) detection of the hybridized nucleic acid fragments. The reagent used in each of these steps and their conditions for use vary depending on the particular application.

In some applications it is necessary to block the hybridization capacity of repetitive sequences. In this case, human genomic DNA is used as an agent to block such hybridization. The preferred size range is from about 200 bp to about 1000 bases,

more preferably between about 400 to about 800 bp for double stranded, nick translated nucleic acids.

Hybridization protocols for the particular applications disclosed here are described in Pinkel *et al.* *Proc. Natl. Acad. Sci. USA*, 85: 9138-9142 (1988) and in EPO Pub. No. 430,402. Suitable hybridization protocols can also be found in *Methods in Molecular Biology Vol. 33: In Situ Hybridization Protocols*, K.H.A. Choo, ed., Humana Press, Totowa, New Jersey, (1994). In a particularly preferred embodiment, the hybridization protocol of Kallioniemi *et al.*, *Proc. Natl Acad Sci USA*, 89: 5321-5325 (1992) is used.

Typically, it is desirable to use dual color FISH, in which two probes are utilized, each labelled by a different fluorescent dye. A test probe that hybridizes to the region of interest is labelled with one dye, and a control probe that hybridizes to a different region is labelled with a second dye. A nucleic acid that hybridizes to a stable portion of the chromosome of interest, such as the centromere region, is often most useful as the control probe. In this way, differences between efficiency of hybridization from sample to sample can be accounted for.

The FISH methods for detecting chromosomal abnormalities can be performed on nanogram quantities of the subject nucleic acids. Paraffin embedded tumor sections can be used, as can fresh or frozen material. Because FISH can be applied to the limited material, touch preparations prepared from uncultured primary tumors can also be used (*see, e.g.*, Kallioniemi, A. *et al.*, *Cytogenet. Cell Genet.* 60: 190-193 (1992)). For instance, small biopsy tissue samples from tumors can be used for touch preparations (*see, e.g.*, Kallioniemi, A. *et al.*, *Cytogenet. Cell Genet.* 60: 190-193 (1992)). Small numbers of cells obtained from aspiration biopsy or cells in bodily fluids (*e.g.*, blood, urine, sputum and the like) can also be analyzed. For prenatal diagnosis, appropriate samples will include amniotic fluid and the like.

#### **Southern Blots**

In a Southern Blot, a genomic or cDNA (typically fragmented and separated on an electrophoretic gel) is hybridized to a probe specific for the target region. Comparison of the intensity of the hybridization signal from the probe for the target region (*e.g.*, 20q13) with the signal from a probe directed to a control (non amplified) such as centromeric DNA, provides an estimate of the relative copy number of the target nucleic acid.

### **Detecting Mutations in Genes from the 20q13 Amplicon**

The cDNA sequences disclosed here can also be used for detecting mutations (*e.g.*, substitutions, insertions, and deletions) within the corresponding endogenous genes. One of skill will recognize that the nucleic acid hybridization techniques generally described above can be adapted to detect such mutations. For instance, oligonucleotide probes that distinguish between mutant and wild-type forms of the target gene can be used in standard hybridization assays. In some embodiments, amplification (*e.g.*, using PCR) can be used to increase copy number of the target sequence prior to hybridization.

### **Assays for detecting 20q13 amplicon proteins.**

As indicated above, this invention identifies protein products of genes in the 20q13 amplicon that are associated with various cancers. In particular, it was shown that 20q13 proteins were overexpressed in various cancers. The presence or absence and/or level of expression of 20q13 proteins can be indicative of the presence, absence, or extent of a cancer. Thus, 20q13 proteins can provide useful diagnostic markers.

The 20q13 amplicon proteins (*e.g.*, ZABC1 or 1b1) can be detected and quantified by any of a number of means well known to those of skill in the art. These may include analytic biochemical methods such as electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC), hyperdiffusion chromatography, and the like, or various immunological methods such as fluid or gel precipitin reactions, immunodiffusion (single or double), immunoelectrophoresis, radioimmunoassay(RIA), enzyme-linked immunosorbent assays (ELISAs), immunofluorescent assays, western blotting, and the like.

In one preferred embodiment, the 20q13 amplicon proteins are detected in an electrophoretic protein separation such as a one dimensional or two-dimensional electrophoresis, while in a most preferred embodiment, the 20q13 amplicon proteins are detected using an immunoassay.

As used herein, an immunoassay is an assay that utilizes an antibody to specifically bind to the analyte (*e.g.*, ZABC1 or 1b1 proteins). The immunoassay is thus characterized by detection of specific binding of a 20q13 amplicon protein to an anti-20q13 amplicon antibody (*e.g.*, anti-ZABC1 or anti-1b1) as opposed to the use of other physical or chemical properties to isolate, target, and quantify the analyte.

The collection of biological sample and subsequent testing for 20q13 amplicon protein(s) is discussed in more detail below.

#### **A) Sample Collection and Processing**

The 20q13 amplicon proteins are preferably quantified in a biological sample derived from a mammal, more preferably from a human patient or from a porcine, murine, feline, canine, or bovine. As used herein, a biological sample is a sample of biological tissue or fluid that contains a 20q13 amplicon protein concentration that may be correlated with a 20q13 amplification. Particularly preferred biological samples include, but are not limited to biological fluids such as blood or urine, or tissue samples including, but not limited to tissue biopsy (*e.g.*, needle biopsy) samples.

The biological sample may be pretreated as necessary by dilution in an appropriate buffer solution or concentrated, if desired. Any of a number of standard aqueous buffer solutions, employing one of a variety of buffers, such as phosphate, Tris, or the like, at physiological pH can be used.

#### **B) Electrophoretic Assays.**

As indicated above, the presence or absence of 20q13 amplicon proteins in a biological tissue may be determined using electrophoretic methods. Means of detecting proteins using electrophoretic techniques are well known to those of skill in the art (*see generally*, R. Scopes (1982) *Protein Purification*, Springer-Verlag, N.Y.; Deutscher, (1990) *Methods in Enzymology Vol. 182: Guide to Protein Purification.*, Academic Press, Inc., N.Y.). In a preferred embodiment, the 20q13 amplicon proteins are detected using one-dimensional or two-dimensional electrophoresis. A particularly preferred two-dimensional electrophoresis separation relies on isoelectric focusing (IEF) in immobilized pH gradients for one dimension and polyacrylamide gels for the second dimension. Such assays are described in the cited references and by Patton *et al.* (1990) *Biotechniques* 8: 518.

#### **C) Immunological Binding Assays.**

In a preferred embodiment, the 20q13 amplicon are detected and/or quantified using any of a number of well recognized immunological binding assays (*see, e.g.*, U.S. Patents 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, see also *Methods in Cell Biology Volume 37: Antibodies in Cell Biology*, Asai, ed. Academic Press, Inc. New York (1993); *Basic and Clinical Immunology* 7th Edition, Stites & Terr, eds. (1991).

Immunological binding assays (or immunoassays) typically utilize a “capture agent” to specifically bind to and often immobilize the analyte (in this case 20q13 amplicon). The capture agent is a moiety that specifically binds to the analyte. In a preferred embodiment, the capture agent is an antibody that specifically binds 20q13 amplicon protein(s).

The antibody (*e.g.*, anti-ZABC1 or anti-1b1) may be produced by any of a number of means well known to those of skill in the art (*see, e.g. Methods in Cell Biology Volume 37: Antibodies in Cell Biology*, Asai, ed. Academic Press, Inc. New York (1993); and *Basic and Clinical Immunology* 7th Edition, Stites & Terr, eds. (1991)). The antibody may be a whole antibody or an antibody fragment. It may be polyclonal or monoclonal, and it may be produced by challenging an organism (*e.g.* mouse, rat, rabbit, *etc.*) with a 20q13 amplicon protein or an epitope derived therefrom. Alternatively, the antibody may be produced *de novo* using recombinant DNA methodology. The antibody can also be selected from a phage display library screened against 20q13 amplicon (*see, e.g. Vaughan et al. (1996) Nature Biotechnology*, 14: 309-314 and references therein).

Immunoassays also often utilize a labeling agent to specifically bind to and label the binding complex formed by the capture agent and the analyte. The labeling agent may itself be one of the moieties comprising the antibody/analyte complex. Thus, the labeling agent may be a labeled 20q13 amplicon protein or a labeled anti-20q13 amplicon antibody. Alternatively, the labeling agent may be a third moiety, such as another antibody, that specifically binds to the antibody/20q13 amplicon protein complex.

In a preferred embodiment, the labeling agent is a second human 20q13 amplicon protein antibody bearing a label. Alternatively, the second 20q13 amplicon protein antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G may also be used as the label agent. These proteins are normal constituents of the cell walls of streptococcal bacteria. They exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a

variety of species. See, generally Kronval, *et al.*, *J. Immunol.*, 111:1401-1406 (1973), and Akerstrom, *et al.*, *J. Immunol.*, 135:2589-2542 (1985).

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, preferably from about 5 minutes to about 24 hours. However, the incubation time will depend upon the assay format, analyte, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

### 1) Non-Competitive Assay Formats.

Immunoassays for detecting 20q13 amplicon proteins may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of captured analyte (in this case 20q13 amplicon) is directly measured. In one preferred "sandwich" assay, for example, the capture agent (anti-20q13 amplicon protein antibodies) can be bound directly to a solid substrate where they are immobilized. These immobilized antibodies then capture 20q13 amplicon protein present in the test sample. The 20q13 amplicon protein thus immobilized is then bound by a labeling agent, such as a second human 20q13 amplicon protein antibody bearing a label. Alternatively, the second 20q13 amplicon protein antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

### 2. Competitive assay formats.

In competitive assays, the amount of analyte (20q13 amplicon protein) present in the sample is measured indirectly by measuring the amount of an added (exogenous) analyte (20q13 amplicon proteins such as ZABC1 or 1b1 protein) displaced (or competed away) from a capture agent (*e.g.*, anti-ZABC1 or anti-1b1 antibody) by the analyte present in the sample. In one competitive assay, a known amount of, in this case, 20q13 amplicon protein is added to the sample and the sample is then contacted with a capture agent, in this case an antibody that specifically binds 20q13 amplicon protein. The amount of 20q13 amplicon protein bound to the antibody is inversely proportional to the concentration of 20q13 amplicon protein present in the sample.

In a particularly preferred embodiment, the anti-20q13 protein antibody is immobilized on a solid substrate. The amount of 20q13 amplicon protein bound to the antibody may be determined either by measuring the amount of 20q13 amplicon present in an 20q13 amplicon protein/antibody complex, or alternatively by measuring the amount of remaining uncomplexed 20q13 amplicon protein. The amount of 20q13 amplicon protein may be detected by providing a labeled 20q13 amplicon protein.

A hapten inhibition assay is another preferred competitive assay. In this assay a known analyte, in this case 20q13 amplicon protein is immobilized on a solid substrate. A known amount of anti-20q13 amplicon protein antibody is added to the sample, and the sample is then contacted with the immobilized 20q13 amplicon protein. In this case, the amount of anti-20q13 amplicon protein antibody bound to the immobilized 20q13 amplicon protein is inversely proportional to the amount of 20q13 amplicon protein present in the sample. Again the amount of immobilized antibody may be detected by detecting either the immobilized fraction of antibody or the fraction of the antibody that remains in solution. Detection may be direct where the antibody is labeled or indirect by the subsequent addition of a labeled moiety that specifically binds to the antibody as described above.

### **3. Other Assay Formats**

In a particularly preferred embodiment, Western blot (immunoblot) analysis is used to detect and quantify the presence of 20q13 amplicon protein in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support, (such as a nitrocellulose filter, a nylon filter, or derivatized nylon filter), and incubating the sample with the antibodies that specifically bind 20q13 amplicon protein. The anti-20q13 amplicon protein antibodies specifically bind to 20q13 amplicon protein on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (e.g., labeled sheep anti-mouse antibodies) that specifically bind to the anti-20q13 amplicon protein.

Other assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (e.g., antibodies) and release encapsulated reagents or markers. The released chemicals are then detected according to standard techniques (see, Monroe *et al.* (1986) *Amer. Clin. Prod. Rev.* 5:34-41).

#### **D) Reduction of Non-Specific Binding.**

One of skill in the art will appreciate that it is often desirable to reduce non-specific binding in immunoassays. Particularly, where the assay involves an antigen or antibody immobilized on a solid substrate it is desirable to minimize the amount of non-specific binding to the substrate. Means of reducing such non-specific binding are well known to those of skill in the art. Typically, this involves coating the substrate with a proteinaceous composition. In particular, protein compositions such as bovine serum albumin (BSA), nonfat powdered milk, and gelatin are widely used with powdered milk being most preferred.

#### **E) Labels.**

The particular label or detectable group used in the assay is not a critical aspect of the invention, so long as it does not significantly interfere with the specific binding of the antibody used in the assay. The detectable group can be any material having a detectable physical or chemical property. Such detectable labels have been well-developed in the field of immunoassays and, in general, most any label useful in such methods can be applied to the present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (*e.g.* Dynabeads™), fluorescent dyes (*e.g.*, fluorescein isothiocyanate, texas red, rhodamine, and the like), radiolabels (*e.g.*,  $^3\text{H}$ ,  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^{32}\text{P}$ ), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.* polystyrene, polypropylene, latex, *etc.*) beads.

The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

Non-radioactive labels are often attached by indirect means. Generally, a ligand molecule (*e.g.*, biotin) is covalently bound to the molecule. The ligand then binds to an anti-ligand (*e.g.*, streptavidin) molecule which is either inherently detectable or covalently bound to a signal system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound. A number of ligands and anti-ligands can be used. Where a ligand has a natural anti-ligand, for example, biotin, thyroxine, and



cortisol, it can be used in conjunction with the labeled, naturally occurring anti-ligands. Alternatively, any haptenic or antigenic compound can be used in combination with an antibody.

The molecules can also be conjugated directly to signal generating compounds, *e.g.*, by conjugation with an enzyme or fluorophore. Enzymes of interest as labels will primarily be hydrolases, particularly phosphatases, esterases and glycosidases, or oxidoreductases, particularly peroxidases. Fluorescent compounds include fluorescein and its derivatives, rhodamine and its derivatives, dansyl, umbelliferone, *etc.* Chemiluminescent compounds include luciferin, and 2,3-dihydrophthalazinediones, *e.g.*, luminol. For a review of various labeling or signal producing systems which may be used, see, U.S. Patent No. 4,391,904).

Means of detecting labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually, by means of photographic film, by the use of electronic detectors such as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product. Finally simple colorimetric labels may be detected simply by observing the color associated with the label. Thus, in various dipstick assays, conjugated gold often appears pink, while various conjugated beads appear the color of the bead.

Some assay formats do not require the use of labeled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need be labeled and the presence of the target antibody is detected by simple visual inspection.

#### **G) Substrates.**

As mentioned above, depending upon the assay, various components, including the antigen, target antibody, or anti-human antibody, may be bound to a solid surface. Many methods for immobilizing biomolecules to a variety of solid surfaces are known in the art. For instance, the solid surface may be a membrane (*e.g.*,

nitrocellulose), a microtiter dish (*e.g.*, PVC, polypropylene, or polystyrene), a test tube (glass or plastic), a dipstick (*e.g.* glass, PVC, polypropylene, polystyrene, latex, and the like), a microcentrifuge tube, or a glass or plastic bead. The desired component may be covalently bound or noncovalently attached through nonspecific bonding.

5                   A wide variety of organic and inorganic polymers, both natural and synthetic may be employed as the material for the solid surface. Illustrative polymers include polyethylene, polypropylene, poly(4-methylbutene), polystyrene, polymethacrylate, poly(ethylene terephthalate), rayon, nylon, poly(vinyl butyrate), polyvinylidene difluoride (PVDF), silicones, polyformaldehyde, cellulose, cellulose  
10 acetate, nitrocellulose, and the like. Other materials which may be employed, include paper, glasses, ceramics, metals, metalloids, semiconductive materials, cements or the like. In addition, are included substances that form gels, such as proteins (*e.g.*, gelatins), lipopolysaccharides, silicates, agarose and polyacrylamides can be used. Polymers which form several aqueous phases, such as dextrans, polyalkylene glycols or  
15 surfactants, such as phospholipids, long chain (12-24 carbon atoms) alkyl ammonium salts and the like are also suitable. Where the solid surface is porous, various pore sizes may be employed depending upon the nature of the system.

                  In preparing the surface, a plurality of different materials may be employed, particularly as laminates, to obtain various properties. For example, protein  
20 coatings, such as gelatin can be used to avoid non-specific binding, simplify covalent conjugation, enhance signal detection or the like.

                  If covalent bonding between a compound and the surface is desired, the surface will usually be polyfunctional or be capable of being polyfunctionalized. Functional groups which may be present on the surface and used for linking can include  
25 carboxylic acids, aldehydes, amino groups, cyano groups, ethylenic groups, hydroxyl groups, mercapto groups and the like. The manner of linking a wide variety of compounds to various surfaces is well known and is amply illustrated in the literature. See, for example, *Immobilized Enzymes*, Ichiro Chibata, Halsted Press, New York, 1978, and Cuatrecasas (1970) *J. Biol. Chem.* 245 3059).

30                   In addition to covalent bonding, various methods for noncovalently binding an assay component can be used. Noncovalent binding is typically nonspecific absorption of a compound to the surface. Typically, the surface is blocked with a second compound to prevent nonspecific binding of labeled assay components. Alternatively,

the surface is designed such that it nonspecifically binds one component but does not significantly bind another. For example, a surface bearing a lectin such as Concanavalin A will bind a carbohydrate containing compound but not a labeled protein that lacks glycosylation. Various solid surfaces for use in noncovalent attachment of assay components are reviewed in U.S. Patent Nos. 4,447,576 and 4,254,082.

#### **Kits Containing 20q13 Amplicon Probes.**

This invention also provides diagnostic kits for the detection of chromosomal abnormalities at 20q13. In a preferred embodiment, the kits include one or more probes to the 20q13 amplicon and/or antibodies to a 20q13 amplicon (*e.g.*, anti-ZABC1 or *anti-Ib1*) described herein. The kits can additionally include blocking probes, instructional materials describing how to use the kit contents in detecting 20q13 amplicons. The kits may also include one or more of the following: various labels or labeling agents to facilitate the detection of the probes, reagents for the hybridization including buffers, a metaphase spread, bovine serum albumin (BSA) and other blocking agents, sampling devices including fine needles, swabs, aspirators and the like, positive and negative hybridization controls and so forth.

#### **Expression of cDNA clones**

One may express the desired polypeptides encoded by the cDNA clones disclosed here, or by subcloning cDNA portions of genomic sequences in a recombinantly engineered cell such as bacteria, yeast, insect (especially employing baculoviral vectors), or mammalian cell. It is expected that those of skill in the art are knowledgeable in the numerous expression systems available for expression of the cDNAs. No attempt to describe in detail the various methods known for the expression of proteins in prokaryotes or eukaryotes will be made.

In brief summary, the expression of natural or synthetic nucleic acids encoding polypeptides of the invention will typically be achieved by operably linking the DNA or cDNA to a promoter (which is either constitutive or inducible), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation terminators, initiation sequences, and promoters useful for regulation of the expression of the DNA encoding the polypeptides. To obtain high level expression of a cloned gene, it is desirable to construct expression plasmids which

contain, at the minimum, a strong promoter to direct transcription, a ribosome binding site for translational initiation, and a transcription/translation terminator.

Examples of regulatory regions suitable for this purpose in *E. coli* are the promoter and operator region of the *E. coli* tryptophan biosynthetic pathway as described by Yanofsky, C., 1984, *J. Bacteriol.*, 158:1018-1024 and the leftward promoter of phage lambda ( $P_L$ ) as described by Herskowitz, I. and Hagen, D., 1980, *Ann. Rev. Genet.*, 14:399-445. The inclusion of selection markers in DNA vectors transformed in *E. coli* is also useful. Examples of such markers include genes specifying resistance to ampicillin, tetracycline, or chloramphenicol. Expression systems are available using *E. coli*, *Bacillus* sp. (Palva, I *et al.*, 1983, *Gene* 22:229-235; Mosbach, K. *et al.* *Nature*, 302:543-545 and *Salmonella*. *E. coli* systems are preferred.

The polypeptides produced by prokaryote cells may not necessarily fold properly. During purification from *E. coli*, the expressed polypeptides may first be denatured and then renatured. This can be accomplished by solubilizing the bacterially produced proteins in a chaotropic agent such as guanidine HCl and reducing all the cysteine residues with a reducing agent such as beta-mercaptoethanol. The polypeptides are then renatured, either by slow dialysis or by gel filtration. U.S. Patent No. 4,511,503.

A variety of eukaryotic expression systems such as yeast, insect cell lines and mammalian cells, are known to those of skill in the art. As explained briefly below, the polypeptides may also be expressed in these eukaryotic systems.

Synthesis of heterologous proteins in yeast is well known and described. *Methods in Yeast Genetics*, Sherman, F., *et al.*, Cold Spring Harbor Laboratory, (1982) is a well recognized work describing the various methods available to produce the polypeptides in yeast. A number of yeast expression plasmids like YEp6, YEp13, YEp4 can be used as vectors. A gene of interest can be fused to any of the promoters in various yeast vectors. The above-mentioned plasmids have been fully described in the literature (Botstein, *et al.*, 1979, *Gene*, 8:17-24; Broach, *et al.*, 1979, *Gene*, 8:121-133).

Illustrative of cell cultures useful for the production of the polypeptides are cells of insect or mammalian origin. Mammalian cell systems often will be in the form of monolayers of cells although mammalian cell suspensions may also be used. Illustrative examples of mammalian cell lines include VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, W138, BHK, Cos-7 or MDCK cell lines.

As indicated above, the vector, *e. g.*, a plasmid, which is used to transform the host cell, preferably contains DNA sequences to initiate transcription and sequences to control the translation of the antigen gene sequence. These sequences are referred to as expression control sequences. When the host cell is of insect or mammalian origin illustrative expression control sequences are often obtained from the SV-40 promoter (Science, 222:524-527, 1983), the CMV I.E. Promoter (Proc. Natl. Acad. Sci. 81:659-663, 1984) or the metallothionein promoter (Nature 296:39-42, 1982). The cloning vector containing the expression control sequences is cleaved using restriction enzymes and adjusted in size as necessary or desirable and ligated with the desired DNA by means well known in the art.

As with yeast, when higher animal host cells are employed, polyadenylation or transcription terminator sequences from known mammalian genes need to be incorporated into the vector. An example of a terminator sequence is the polyadenylation sequence from the bovine growth hormone gene. Sequences for accurate splicing of the transcript may also be included. An example of a splicing sequence is the VPI intron from SV40 (Sprague, J. *et al.*, 1983, J. Virol. 45: 773-781).

Additionally, gene sequences to control replication in the host cell may be incorporated into the vector such as those found in bovine papilloma virus type-vectors. Saveria-Campo, M., 1985, "Bovine Papilloma virus DNA a Eukaryotic Cloning Vector" in DNA Cloning Vol. II a Practical Approach Ed. D.M. Glover, IRL Press, Arlington, Virginia pp. 213-238.

#### **Therapeutic and other uses of cDNAs and their gene products**

The cDNA sequences and the polypeptide products of the invention can be used to modulate the activity of the gene products of the endogenous genes corresponding to the cDNAs. By modulating activity of the gene products, pathological conditions associated with their expression or lack of expression can be treated. Any of a number of techniques well known to those of skill in the art can be used for this purpose.

The cDNAs of the invention are particularly used for the treatment of various cancers such as cancers of the breast, ovary, bladder, head and neck, and colon. Other diseases may also be treated with the sequences of the invention. For instance, as noted above, GCAP (SEQ. ID. No. 6) encodes a guanine cyclase activating protein which is involved in the biosynthesis of cyclic AMP. Mutations in genes involved in the biosynthesis of cyclic AMP are known to be associated with hereditary retinal

degenerative diseases. These diseases are a group of inherited conditions in which progressive, bilateral degeneration of retinal structures leads to loss of retinal function. These diseases include age-related macular degeneration, a leading cause of visual impairment in the elderly; Leber's congenital amaurosis, which causes its victims to be born blind; and retinitis pigmentosa ("RP"), one of the most common forms of inherited blindness. RP is the name given to those inherited retinopathies which are characterized by loss of retinal photoreceptors (rods and cones), with retinal electrical responses to light flashes (i.e. electroretinograms, or "ERGs") that are reduced in amplitude.

The mechanism of retinal photoreceptor loss or cell death in different retinal degenerations is not fully understood. Mutations in a number of different genes have been identified as the primary genetic lesion in different forms of human RP. Affected genes include rhodopsin, the alpha and beta subunits of cGMP photodiesterase, and peripherin-RDS (Dryja, T. P. et al., *Invest. Ophthalmol. Vis. Sci.* 36, 1197-1200 (1995)). In all cases the manifestations of the disorder regardless of the specific primary genetic mutation is similar, resulting in photoreceptor cell degeneration and blindness.

Studies on animal models of retinal degeneration have been the focus of many laboratories during the last decade. The mechanisms that are altered in some of the mutations leading to blindness have been elucidated. This would include the inherited disorders of the *rd* mouse. The *rd* gene encodes the beta subunit of cGMP-phosphodiesterase (PDE) (Bowes, C. et al., *Nature* 347, 677-680 (1990)), an enzyme of fundamental importance in normal visual function because it is a key component in the cascade of events that takes place in phototransduction.

The polypeptides encoded by the cDNAs of the invention can be used as immunogens to raise antibodies either polyclonal or monoclonal. The antibodies can be used to detect the polypeptides for diagnostic purposes, as therapeutic agents to inhibit the polypeptides, or as targeting moieties in immunotoxins. The production of monoclonal antibodies against a desired antigen is well known to those of skill in the art and is not reviewed in detail here.

Those skilled in the art recognize that there are many methods for production and manipulation of various immunoglobulin molecules. As used herein, the terms "immunoglobulin" and "antibody" refer to a protein consisting of one or more polypeptides substantially encoded by immunoglobulin genes. Immunoglobulins may

exist in a variety of forms besides antibodies, including for example, Fv, Fab, and F(ab)<sub>2</sub>, as well as in single chains. To raise monoclonal antibodies, antibody-producing cells obtained from immunized animals (*e.g.*, mice) are immortalized and screened, or screened first for the production of the desired antibody and then immortalized. For a discussion of general procedures of monoclonal antibody production see Harlow and Lane, *Antibodies, A Laboratory Manual* Cold Spring Harbor Publications, N.Y. (1988).

The antibodies raised by these techniques can be used in immunodiagnostic assays to detect or quantify the expression of gene products from the nucleic acids disclosed here. For instance, labeled monoclonal antibodies to polypeptides of the invention can be used to detect expression levels in a biological sample. For a review of the general procedures in diagnostic immunoassays, see *Basic and Clinical Immunology* 7th Edition D. Stites and A. Terr ed. (1991).

The polynucleotides of the invention are particularly useful for gene therapy techniques well known to those skilled in the art. Gene therapy as used herein refers to the multitude of techniques by which gene expression may be altered in cells. Such methods include, for instance, introduction of DNA encoding ribozymes or antisense nucleic acids to inhibit expression as well as introduction of functional wild-type genes to replace mutant genes (*e.g.*, using wild-type GCAP genes to treat retinal degeneration). A number of suitable viral vectors are known. Such vectors include retroviral vectors (see Miller, *Curr. Top. Microbiol. Immunol.* 158: 1-24 (1992); Salmons and Gunzburg, *Human Gene Therapy* 4: 129-141 (1993); Miller et al., *Methods in Enzymology* 217: 581-599, (1994)) and adeno-associated vectors (reviewed in Carter, *Curr. Opinion Biotech.* 3: 533-539 (1992); Muzyczka, *Curr. Top. Microbiol. Immunol.* 158: 97-129 (1992)). Other viral vectors that may be used within the methods include adenoviral vectors, herpes viral vectors and Sindbis viral vectors, as generally described in, *e.g.*, Jolly, *Cancer Gene Therapy* 1:51-64 (1994); Latchman, *Molec. Biotechnol.* 2:179-195 (1994); and Johanning et al., *Nucl. Acids Res.* 23:1495-1501 (1995).

Delivery of nucleic acids linked to a heterologous promoter-enhancer element via liposomes is also known (see, *e.g.*, Brigham, *et al.* (1989) *Am. J. Med. Sci.*, 298:278-281; Nabel, *et al.* (1990) *Science*, 249:1285-1288; Hazinski, *et al.* (1991) *Am. J. Resp. Cell Molec. Biol.*, 4:206-209; and Wang and Huang (1987) *Proc. Natl. Acad. Sci. (USA)*, 84:7851-7855); coupled to ligand-specific, cation-based transport systems (Wu and Wu (1988) *J. Biol. Chem.*, 263:14621-14624). Naked DNA expression

vectors have also been described (Nabel *et al.* (1990), *supra*); Wolff *et al.* (1990) *Science*, 247:1465-1468).

The nucleic acids and encoded polypeptides of the invention can be used directedly to inhibit the endogenous genes or their gene products. For instance,

5 Inhibitory nucleic acids may be used to specifically bind to a complementary nucleic acid sequence. By binding to the appropriate target sequence, an RNA-RNA, a DNA-DNA, or RNA-DNA duplex is formed. These nucleic acids are often termed "antisense" because they are usually complementary to the sense or coding strand of the gene, although approaches for use of "sense" nucleic acids have also been developed. The

10 term "inhibitory nucleic acids" as used herein, refers to both "sense" and "antisense" nucleic acids. Inhibitory nucleic acid methods encompass a number of different approaches to altering expression of specific genes that operate by different mechanisms.

In brief, inhibitory nucleic acid therapy approaches can be classified into those that target DNA sequences, those that target RNA sequences (including pre-mRNA

15 and mRNA), those that target proteins (sense strand approaches), and those that cause cleavage or chemical modification of the target nucleic acids (ribozymes). These different types of inhibitory nucleic acid technology are described, for instance, in Helene, C. and Toulme, J. (1990) *Biochim. Biophys. Acta.*, 1049:99-125. Inhibitory nucleic acid complementary to regions of c-myc mRNA has been shown to inhibit c-myc

20 protein expression in a human promyelocytic leukemia cell line, HL60, which overexpresses the c-myc protooncogene. See Wickstrom E.L., *et al.*, (1988) *PNAS (USA)*, 85:1028-1032 and Harel-Bellan, A., *et al.*, (1988) *Exp. Med.*, 168:2309-2318.

The encoded polypeptides of the invention can also be used to design molecules (peptidic or nonpeptidic) that inhibit the endogenous proteins by, for instance,

25 inhibiting interaction between the protein and a second molecule specifically recognized by the protein. Methods for designing such molecules are well known to those skilled in the art.

For instance, polypeptides can be designed which have sequence identity with the encoded proteins or may comprise modifications (conservative or non-

30 conservative) of the sequences. The modifications can be selected, for example, to alter their *in vivo* stability. For instance, inclusion of one or more D-amino acids in the peptide typically increases stability, particularly if the D-amino acid residues are substituted at one or both termini of the peptide sequence.



The polypeptides can also be modified by linkage to other molecules. For example, different N- or C-terminal groups may be introduced to alter the molecule's physical and/or chemical properties. Such alterations may be utilized to affect, for example, adhesion, stability, bio-availability, localization or detection of the molecules.

5 For diagnostic purposes, a wide variety of labels may be linked to the terminus, which may provide, directly or indirectly, a detectable signal. Thus, the polypeptides may be modified in a variety of ways for a variety of end purposes while still retaining biological activity.

### EXAMPLES

10 The following examples are offered to illustrate, but not to limit the present invention.

#### Example 1

#### PROGNOSTIC IMPLICATIONS OF AMPLIFICATION OF CHROMOSOMAL REGION 20q13 IN BREAST CANCER

##### 15 Patients and tumor material.

Tumor samples were obtained from 152 women who underwent surgery for breast cancer between 1987 and 1992 at the Tampere University or City Hospitals. One hundred and forty-two samples were from primary breast carcinomas and 11 from metastatic tumors. Specimens from both the primary tumor and a local metastasis were

20 available from one patient. Ten of the primary tumors that were either *in situ* or mucinous carcinomas were excluded from the material, since the specimens were considered inadequate for FISH studies. Of the remaining 132 primary tumors, 128 were invasion ductal and 4 lobular carcinomas. The age of the patients ranged from 29 to 92 years (mean 61). Clinical follow-up was available from 129 patients. Median

25 follow-up period was 45 months (range 1.4-177 months). Radiation therapy was given to 77 of the 129 patients (51 patients with positive and 26 with negative lymph nodes), and systemic adjuvant therapy to 36 patients (33 with endocrine and 3 with cytotoxic chemotherapy). Primary tumor size and axillary node involvement were determined according to the tumor-node metastasis (TNM) classification. The histopathological

30 diagnosis was evaluated according to the World Health Organization (11). The carcinomas were graded on the basis of the tubular arrangement of cancer cells, nuclear atypia, and frequency of mitotic or hyperchromatic nuclear figures according to Bloom and Richardson, *Br. J. Cancer*, 11: 359-377 (1957).

Surgical biopsy specimens were frozen at  $-70^{\circ}\text{C}$  within 15 minutes of removal. Cryostat sections ( $5\text{--}6\text{ }\mu\text{m}$ ) were prepared for intraoperative histopathological diagnosis, and additional thin sections were cut for immunohistochemical studies. One adjacent  $200\text{ }\mu\text{m}$  thick section was cut for DNA flow cytometric and FISH studies.

## 5 Cell preparation for FISH.

After histological verification that the biopsy specimens contained a high proportion of tumor cells, nuclei were isolated from  $200\text{ }\mu\text{m}$  frozen sections according to a modified Vindelov procedure for DNA flow cytometry, fixed and dropped on slides for FISH analysis as described by Hyytinen *et al.*, *Cytometry* 16: 93-99 (1994). Foreskin  
10 fibroblasts were used as negative controls in amplification studies and were prepared by harvesting cells at confluency to obtain G1 phase enriched interphase nuclei. All samples were fixed in methanol-acetic-acid (3:1).

## Probes.

Five probes mapping to the 20q13 region were used ( *see* Stokke, *et al.*,  
15 *Genomics*, 26: 134-137 (1995)). The probes included Pl-clones for melanocortin-3-receptor (probe MC3R, fractional length from p-arm telomere (Flpter 0.81) and phosphoenolpyruvate carboxy kinase (PCK, Flpter 0.84), as well as anonymous cosmid clones RMC20C026 (Flpter 0.79). In addition, RMC20C001 (Flpter 0.825) and RMC20C030 (Flpter 0.85) were used. Probe RMC20C001 was previously  
20 shown to define the region of maximum amplification (Tanner *et al.*, *Cancer Res*, 54: 4257-4260 (1994)). One cosmid probe mapping to the proximal p-arm, RMC20C038 (FLpter 0.237) was used as a chromosome-specific reference probe. Test probes were labeled with biotin-14-dATP and the reference probe with digoxigenin-11-dUTP using nick translation (Kallioniemi *et al.*, *Proc. Natl Acad Sci USA*, 89: 5321-5325 (1992)).

## 25 Fluorescence in situ hybridization.

Two-color FISH was performed using biotin-labeled 20q13-specific probes and digoxigenin-labelled 20p reference probe essentially as described (*Id.*). Tumor samples were postfixed in 4% paraformaldehyde/phosphate-buffered saline for 5 min at  
30 4 C prior to hybridization, dehydrated in 70%, 85% and 100% ethanol, air dried, and incubated for 30 min at  $80^{\circ}\text{C}$ . Slides were denatured in a 70% formamide/2x standard saline citrate solution at  $72\text{--}74^{\circ}\text{C}$  for 3 min, followed by a proteinase K digestion ( $0.5\text{ }\mu\text{g/ml}$ ). The hybridization mixture contained 18 ng of each of the labeled probes and  $10\text{ }\mu\text{g}$  human placental DNA. After hybridization, the probes were detected

immunochemically with avidin-FITC and anti-digoxigenin Rhodamine. Slides were counterstained with 0.2  $\mu$ M 4,6-diamidino-2-phenylindole (DAPI) in an antifade solution.

#### **Fluorescence microscopy and scoring of signals in interphase nuclei.**

A Nikon fluorescence microscope equipped with double band-pass filters (Chromatechnology, Brattleboro, Vermont, USA) and 63 x objective (NA 1.3) was used for simultaneous visualization of FITC and Rhodamine signals. At least 50 non-overlapping nuclei with intact morphology based on the DAPI counterstaining were scored to determine the number of test and reference probe hybridization signals. Leukocytes infiltrating the tumor were excluded from analysis. Control hybridizations to normal fibroblast interphase nuclei were done to ascertain that the probes recognized a single copy target and that the hybridization efficiencies of the test and reference probes were similar.

The scoring results were expressed both as the mean number of hybridization signals per cell and as mean level of amplification (= mean of number of signals relative to the number of reference probe signals).

#### **DNA flow cytometry and steroid receptor analyses.**

DNA flow cytometry was performed from frozen 200  $\mu$ m sections as described by Kallioniemi, *Cytometry* 9: 164-169 (1988). Analysis was carried out using an EPICS C flow cytometer (Coulter Electronics Inc., Hialeah, Florida, USA) and the MultiCycle program (Phoenix Flow Systems, San Diego, California, USA). DNA-index over 1.07 (in over 20% of cells) was used as a criterion for DNA aneuploidy. In DNA aneuploid histograms, the S-phase was analyzed only from the aneuploid clone. Cell cycle evaluation was successful in 86% (108/126) of the tumors.

Estrogen (ER) and progesterone (PR) receptors were detected immunohistochemically from cryostat sections as previously described (17). The staining results were semiquantitatively evaluated and a histoscore greater than or equal to 100 was considered positive for both ER and PR (17).

#### **Statistical Methods.**

Contingency tables were analyzed with Chi square test for trend. Association between S-phase fraction (continuous variable) and 20q13 amplification was analyzed with Kruskal-Wallis test. Analysis of disease-free survival was performed using the BMDPIL program and Mautel-Cox test and Cox's proportional hazards model (BMDP2L program) was used in multivariate regression analysis (Dixon *BMDP*

*Statistical Software*. London, Berkeley, Los Angeles: University of California Press, (1981)).

**Amplification of 20q13 in primary breast carcinomas by fluorescence *in situ* hybridization.**

5                   The minimal region probe RMC20C001 was used in FISH analysis to assess the 20q13 amplification. FISH was used to analyze both the total number of signals in individual tumor cells and to determine the mean level of amplification (mean copy number with the RMC20C001 probe relative to a 20p-reference probe). In addition, the distribution of the number of signals in the tumor nuclei was also assessed.

10 Tumors were classified into three categories: no, low and high level of amplification. Tumors classified as not amplified showed less than 1.5 fold-copy number of the RMC20C001 as compared to the p-arm control. Those classified as having low-level amplification had 1.5-3-fold average level of amplification. Tumors showing over 3-fold average level of amplification were classified as highly amplified.

15                   The highly amplified tumors often showed extensive intratumor heterogeneity with up to 40 signals in individual tumor cells. In highly amplified tumors, the RMC20C001 probe signals were always arranged in clusters by FISH, which indicates location of the amplified DNA sequences in close proximity to one another *e.g.* in a tandem array. Low level 20q13 amplification was found in 29 of the 132 primary

20 tumors (22%), whereas nine cases (6.8%) showed high level amplification. The overall prevalence of increased copy number in 20q13 was thus 29% (38/132).

**Defining the minimal region of amplification.**

25                   The average copy number of four probes flanking RMC20C001 was determined in the nine highly amplified tumors. The flanking probes tested were malanocortin-3-receptor (MC3R, FLpter 0.81), phosphoenolpyruvate carboxykinase (PCK, 0.84), RMC20C026 (0.79) and RMC20C030 (0.85). The amplicon size and location varied slightly from one tumor to another but RMC20C001 was the only probe consistently highly amplified in all nine cases.

**Association of 20q13 amplification with pathological and biological features.**

30                   The 20q13 amplification was significantly associated with high histologic grade of the tumors ( $p=0.01$ ). This correlation was seen both in moderately and highly amplified tumors (Table 4). Amplification of 20q13 was also significantly associated with aneuploidy as determined by DNA flow cytometry ( $p=0.01$ , Table 4) The mean

cell proliferation activity, measured as the percentage of cells in the S-phase fraction, increased ( $p=0.0085$  by Kruskal-Wallis test) with the level of amplification in tumors with no, low and high levels of amplification (Table 4). No association was found with the age of the patient, primary tumor size, axillary nodal or steroid hormone-receptor status (Table 4).

Table 4. Clinicopathological correlations of amplification at chromosomal region 20q13 in 132 primary breast cancers.

Pathobiologic feature	20q13 amplification status			p-value <sup>1</sup>
	NO Number of patients (%)	LOW LEVEL Number of patients (%)	HIGH LEVEL Number of patients (%)	
All primary tumors	94 (71%)	29 (22%)	9 (6.8%)	
Age of patients				
< 50 years	17 (65%)	6 (23%)	3 (12%)	.39
≥ 50 years	77 (73%)	23 (22%)	6 (5.7%)	
Tumor size				
< 2 cm	33 (79%)	7 (17%)	2 (4.8%)	.16
≥ 2 cm	58 (67%)	22 (25%)	7 (8.0%)	
Nodal status				
Negative	49 (67%)	19 (26%)	5 (6.8%)	.41
Positive	41 (75%)	10 (18%)	4 (7.3%)	
Histologic grade				
I - II	72 (76%)	18 (19%)	5 (5.3%)	.01
III	16 (52%)	11 (35%)	4 (13%)	
Estrogen receptor status				
Negative	30 (67%)	10 (22%)	5 (11%)	.42
Positive	59 (72%)	19 (23%)	4 (4.9%)	
Progesterone receptor status				
Negative	57 (69%)	20 (24%)	6 (7.2%)	.53
Positive	32 (74%)	8 (19%)	3 (7.0%)	
DNA ploidy				
Diploid	45 (82%)	8 (14.5%)	2 (3.6%)	.01
Aneuploid	44 (62%)	20 (28%)	7 (10%)	
S-phase fraction (%)	<u>mean ± SD</u> 9.9 ± 7.2	<u>mean ± SD</u> 12.6 ± 6.7	<u>mean ± SD</u> 19.0 ± 10.5	.0085 <sup>1</sup>

<sup>1</sup>. Kruskal-Wallis Test.

#### **Relationship between 20q13 amplification and disease-free survival.**

Disease-free survival of patients with high-level 20q13 amplification was significantly shorter than for patients with no or only low-level amplification ( $p=0.04$ ). Disease-free survival of patients with moderately amplified tumors did not differ significantly from that of patients with no amplification. Among the node-negative

patients (n=79), high level 20q13 amplification was a highly significant prognostic factor for shorter disease-free survival ( $p=0.002$ ), even in multivariate Cox's regression analysis ( $p=0.026$ ) after adjustment for tumor size ER, PR grade, ploidy and S-phase fraction.

#### 5 **20q13 amplification in metastatic breast tumors.**

Two of 11 metastatic breast tumors had low level and one high level 20q13 amplification. Thus, the overall prevalence (27%) of increased 20q13 copy number in metastatic tumors was a similar to that observed in the primary tumors. Both a primary and a metastatic tumor specimens were available from one of the patients.

10 This 29-year old patient developed a pectoral muscle infiltrating metastasis eight months after total mastectomy. The patient did not receive adjuvant or radiation therapy after mastectomy. The majority of tumor cells in the primary tumor showed a low level amplification, although individual tumor cells (less than 5% of total) contained 8-20 copies per cell by FISH. In contrast, all tumor cells from metastasis showed high level  
15 20q13 amplification (12-50 copies per cell). The absolute copy number of the reference probe remained the same suggesting that high level amplification was not a result of an increased degree of aneuploidy.

#### **Diagnostic and Prognostic Value of the 20q13 Amplification.**

The present findings suggest that the newly-discovered 20q13 amplification  
20 may be an important component of the genetic progression pathway of certain breast carcinomas. Specifically, the foregoing experiments establish that: 1) High-level 20q13 amplification, detected in 7% of the tumors, was significantly associated with decreased disease-free survival in node-negative breast cancer patients, as well as with indirect indicators of high-malignant potential, such as high grade and S-phase fraction. 2)  
25 Low-level amplification, which was much more common, was also associated with clinicopathological features of aggressive tumors, but was not prognostically significant. 3) The level of amplification of RMC20C001 remains higher than amplification of nearby candidate genes and loci indicating that a novel oncogene is located in the vicinity of RMC20C001.

30 High-level 20q13 amplification was defined by the presence of more than 3-fold higher copy number of the 20q13 amplification is somewhat lower than the amplification frequencies reported for some of the other breast cancer oncogenes, such as ERBB2 (17q12) and Cyclin-D (11q13) (Borg *et al.*, *Oncogene*, 6: 137-143 (1991), Van

de Vijver *et al. Adv. Canc. Res.*, 61: 25-56 (1993)). However, similar to what has been previously found with these other oncogenes (Swab, *et al., Genes Chrom. Canc.*, 1: 181-193 (1990), Borg *et al., supra.*), high-level 20q13 amplification was more common in tumors with high grade or high S-phase fraction and in cases with poor prognosis.

5 Although only a small number of node-negative patients was analyzed, our results suggest that 20q13 amplification might have independent role as a prognostic indicator. Studies to address this question in large patient materials are warranted. Moreover, based on these survival correlations, the currently unknown, putative oncogene amplified in this locus may confer an aggressive phenotype. Thus, cloning of this gene is an  
10 important goal. Based on the association of amplification with highly proliferative tumors one could hypothesize a role for this gene in the growth regulation of the cell.

The role of the low-level 20q13 amplification as a significant event in tumor progression appears less clear. Low-level amplification was defined as 1.5-3-fold increased average copy number of the 20q13 probe relative to the p-arm control. In  
15 addition, these tumors characteristically lacked individual tumor cells with very high copy numbers, and showed a scattered, not clustered, appearance of the signals. Accurate distinction between high and low level 20q13 amplification can only be reliably done by FISH, whereas Southern and slot blot analyses are likely to be able to detect only high-level amplification, in which substantial elevation of the average gene copy  
20 number takes place. This distinction is important, because only the high amplified tumors were associated with adverse clinical outcome. Tumors with low-level 20q13 amplification appeared to have many clinicopathological features that were in between of those found for tumors with no and those with high level amplification. For example, the average tumor S-phase fraction was lowest in the non-amplified tumors and highest in  
25 the highly amplified tumors. One possibility is that low-level amplification precedes the development of high level amplification. This has been shown to be the case, *e.g.*, in the development of drug resistance-gene amplification *in vitro* (Stark, *Adv. Canc. Res.*, 61: 87-113 (1993)). Evidence supporting this hypothesis was found in one of our patients, whose local metastasis contained a much higher level of 20q13 amplification  
30 than the primary tumor operated 8 months earlier.

Finally, our previous paper reported a 1.5 Mb critical region defined by RMC20C001 probe and exclusion of candidate genes in breast cancer cell lines and in a limited number of primary breast tumors. Results of the present study confirm these

findings by showing conclusively in a larger set of primary tumors that the critical region of amplification is indeed defined by this probe.

5 The present data thus suggest that the high-level 20q13 amplification may be a significant step in the progression of certain breast tumors to a more malignant phenotype. The clinical and prognostic implications of 20q13 amplification are striking and location of the minimal region of amplification at 20q13 has now been defined.

10 It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference for all purposes.

#### Discussion of the Accompanying Sequence Listing

15 SEQ ID NOs:1-10 and 12-13 provide nucleic acid sequences. In each case, the information is presented as a DNA sequence. One of skill will readily understand that the sequence also describes the corresponding RNA (*i.e.*, by substitution of the T residues with U residues) and a variety of conservatively modified variations thereof. The complementary sequence is fully described by comparison to the existing  
20 sequence, *i.e.*, the complementary sequence is obtained by using standard base pairing rules for DNA (*e.g.*, A to T, C to G). In addition, the nucleic acid sequence provides the corresponding amino acid sequence by translating the given DNA sequence using the genetic code.

25 For SEQ ID NO 11, the information is presented as a polypeptide sequence. One of skill will readily understand that the sequence also describes all of the corresponding RNA and DNA sequences which encode the polypeptide, by conversion of the amino acid sequence into the corresponding nucleotide sequence using the genetic code, by alternately assigning each possible codon in each possible codon position. Similarly, each nucleic acid sequence which is provided also inherently provides all of the  
30 nucleic acids which encode the same protein, since one of skill simply translates a selected nucleic acid into a protein and then uses the genetic code to reverse translate all possible nucleic acids from the amino acid sequence.



The sequences also provide a variety of conservatively modified variations by substituting appropriate residues with the exemplar conservative amino acid substitutions provided, *e.g.*, in the Definitions section above.

## SEQUENCE LISTING

SEQ. ID. No. 1

3bf4 3000 bp

5 CCGCCGGCCGGGCGCCTGGCTGCACTCAGCGCCGGAGCCGGGAGCTAGCGGCCGCCCATGTCCCACCAGAC  
CGGCATCCAAGCAAGTGAAGATGTTAAAGAGATCTTTGCCAGAGCCAGAAATGGAAAGTACAGACTTCTGAAAA  
TATCTATTGAAATGAGCAACTTGTGATTGGATCATATAGTCAGCCTTCAGATTCTGGGATAAGGATTATGAT  
TCCTTTGTTTTACCCCTGTTGGAGGACAAACAACCATGCTATATATTATTCAGGTTAGATTCTCAGAATGCCCCA  
GGGATATGAATGGATATTTCATTGTCATGGTCTCCAGATCATTCTCATGTTCTGCAAAAAATGTTGTATGCAGCAA  
10 CAAGAGCAACTCTGAAGAAGGAATTTGGAGGTGGCCACATTAAAGATGAAGTATTTGGAAACAGTAAAGGAAGAT  
GTATCATTACATGGATATAAAAAATACTTGCTGTCACAATCTTCCCCTGCCCCACTGACTGCAGCTGAGGAAGA  
ACTACGACAGATTAAAAATCAATGAGGTACAGACTGACGTGGGTGTGGACACTAAGCATCAAACTACAAGGAG  
TAGCATTTCCTCCATTCTCGAGAAGCCTTTCAGGCTTTGGAAAAAATGAATAATAGACAGCTCAACTATGTGCAG  
TTGGAAATAGATATAAAAAATGAAATTATAATTTTGGCCAAACAACAAATACAGAACTGAAAGATTTGCCAAA  
15 GAGGATTCCCAAGGATTGAGCTCGTTACCATTTCTTCTGTATAAACATTCCCATGAAGGAGACTATTTAGAGT  
CCATAGTTTTTATTTATTCATGCTGGGATACACATGCAGTATAAGAGAGCGGATGCTGTATTCTAGCTGCAG  
AGCCGTCTGCTAGAAATTTGTAGAAAGACAACACAAATGGATGTAATTAGAAAGATCGAGATAGACAATGGGGA  
TGAGTTGACTGCAGACTTCCTTTTATGAAGAAGTACATCCCAAGCAGCATGCACACAAGCAAAGTTTGGCAAAAC  
CAAAAGGTCTGCAGGAAAAAGAGGAATTCGAAGACTAATTAGGGGCCAGCGGAACTGAAGCTACTACTGAT  
20 TAAAGTCATCACATTAAACATTGTAATACAGTCTTTTAAAAAGTCCAGCTTTTAGTACAGGAGAACTGAAATCA  
TTCCATGTTGATATAAAGTAGGGAAAAAATTTGACTTTTGGAAAAATAGCACTTTTCACTTCTGTGTGTTTTT  
AAAATTAATGTTTATAGAAGACTCATGATTTCTATTTTGGAGTTAAAGCTAGAAAAGGGTTCAACATAATGTTTA  
ATTTTGTACACTGTTTTTATAGCGTTGATTCCACACTTCAAATACTTCTTAAATTTTATACAGTTGGGCCAG  
TTCTAGAAAGTCTGATGCTCAAAGGGTAACTTACTACTTTCTTGTGGGACAGAAAGACCTTAAATATTCAT  
25 ATTACTTAATGAATATGTTAAGGACCAGGCTAGAGTATTTCTAAGCTGGAACTTAGTGTGCCTTGGAAAAGC  
CGCAAGTTGCTTACTCCGAGTAGCTGTGCTAGCTCTGTCAGACTGTAGGATCATGTCTGCAACTTTTAGAAATA  
GTGCTTTATATTGCAGCAGTCTTTTATATTGACTTTTTTTTTTAAATAGCATTAAAAATTCAGATCAGCTCACTCT  
GAAACTTTAAGGGTACCAGATATTTTCTATACTGCAGGATTTCTGATGACATTGAAAGACTTTAAACAGCCTTA  
GTAAATTATCTTTCTAATGCTCTGTGAGGCCAAACATTTATGTTTCAGATTGAAATTTAAATTAATATCATTCAA  
30 AAGGAAACAAAAATGTTGAGTTTTAAAAATCAGGATTGACTTTTTTCTCCAAAACCATACATTTATGGGCAAA  
TTGTGTTCTTTATCACTTCCGAGCAATACTCAGATTTAAAAATTACTTTAAAGTCTGGTACTTAACAGGCTAA  
CGTAGATAAAACACCTTAATAATCTCAGTTAATACTGTATTTCAAACACATTTAACTGTTTTCTAATGCTTTGC  
ATTATCAGTTACAACCTAGAGAGATTTTGGCCTCATATTTCTTTGATCTTGAATAGAGGGAGCTAGAACAC  
TTAATGTTTAACTGTTAAACCTGCTGCAAGAGCCATAACTTTGAGGCATTTCTAAATGAACTGTGGGGATCC  
35 AGGATTTGTAATTTCTTGATCTAACTTTATGCTGCATAAACTCACTTATCGGAAATGCACATTTCATAGTGTGA  
AGCACTCATTCTAAACCTTATTATCTAAGGTAATATGTCACCTTTCAGAAATTTGTGTTTCGAGTAAGTAAAG  
CATATTAGAATAATTGTGGGTGACAGATTTTTTAAATAGAATTTAGAGTATTTGGGGTTTTGTTTGTTTACAA  
ATAATCAGACTATAATATTTAAACATGCAAAATAACTGACAATAATGTTGCACTTGTCTACTAAAGATATAAGT  
TGTTCCATGGGTGTACACGTAGACAGACACACATACCCCAAATTATTGCATTAGAATCCTGGAGCAGACCAT  
40 AGCTGAAGCTGTTATTTTCAGTCAGGAAGACTACCTGTCTAAGGTATAAAATAATTTAGAAGTGAATGTTTT  
TCTGTACCCTATGTGCAATTATACTCTAAATTCCTACACTACATTAAAGTAAATGGACATTCCAGAATAT  
AGATGTGATTATAGTCTTAACTAATTATTATTAAACCAATGATTGCTGAAAATCAGTGATGCATTTGTTATAG  
AGTATAACTCATCGTTTACAGTATGTTTTAGTTGGCAGTATCATACCTAGATGGTGAATAACATATTTCCAGTA  
AATTTATATAGCAGTGAAGAATTACATGCCTTCTGGTGGACATTTTATAAGTGCATTTTATATCAATAAAAA  
45 TTTTTCTCTTTAAAAAACAAGAAAAA

SEQ. ID. No. 2

1b11 723 bp

50 TGGAAGCTGTCACTGGTTACCGTCTCTAACGTTGGACTCTTAAGAAAATGATTATTCCTGGTTTTCTAGACAGGCC  
AAATGTAATTCACCTACGTGGCAGATTAAAGAGGTGGGCTTACTAGATTTGATTGGGTATTGAGCATGCTCTGA  
ATGACAGTCCCCAAAAAGGACCTCTTATCCGTTCTTCCCCTTGGGGAAGGGCTTTTGCCACTTCCATGTCAATG  
TGCCAGTTGAGCTTGAAATTTGGTGCCTGTGACAACATAAGCATTACTTCTCCAAGATGTGCTGTGTAGAAAT  
GGTCATAGATTCAAACTGTAGCTACTATGTGGACAGGGGGCAGCAAGGACCCACTTTGTAAAAACATGTTTT  
55 GGGGGAATGTTTTGTTTTTCAATTTCTTATTACCTGGCAAAATAATCCAGGTGGTGTGTGAGTCAACAGTAGAG  
ATTATAAAGTCCAAGGAAGTAGAATCAGCCTTACAAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACC  
CACWGAAGCGAACTCAAAAGAGAAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTTCTCAGACAAATGGTAA  
GCCCCTTACTTCCAGTATAGGAAACCTAAGATACCTAGAGCGGCTTTTGGGAACAATGGGCTCATGCCACAGGT  
AGTAGGAGACATAATTGTAGCTGGTGTGTATGGAATGTGAATGGAATATGGATTGCG

SEQ. ID. No. 3

cc49 1507 bp

60 GCAGGTTGCTGGGATTGACTTCTTGCTCAATTGAAACACTCATTCAATGGAGACAAAGAGCACTAATGCTTTGT  
GCTGATTCATATTGAAATCGAGGCATTGGGAACCTGTATGCCTTGTGTTGTGGAAAGAACCAGTGACACCATCA  
CTGAGCTTCTTAAAGTTCGAAGAAGTTAGAGGACTATACATTTCTTTTGAACTTTTATAATAAATATTTGCT  
CTGTTTTTGAACCCAGGACTGTTAGAGGTGAGTGACAGGTCTTACAGTGGCCTTAATCCAACCTCCAGAAATT

GCCCAACGGAACCTTTGAGATTATATGCAATCGAAAGTGACAGGAAACATGCCAACTCAATCCCTCTTAATGTAC  
 ATGGATGGCCAAGAGTGATTGGCAGCTCTCTTGCCAGTCCGATGGAGATGGAGATGCCTTGTCAATGAAAGGGG  
 CCNCTGTTGTCAATTCCGAGCTACACAAAGAAAAAATGTCAATCCGAATCGAGGGGAATATGCCCTTGGATTG  
 CATGTTCTGCAGCCAGACCTTACACATTGAGAAGACCTTAATAAACATGTCTTAATGCAACACCGGCCTACCC  
 TCTGTGAACAGCAGTTCTTTCGGGTTGAAGCAGAGTATCTCAGTCCGCTTGATAAAAGTCAAGTGCAGAACAGAA  
 CCTCCCAAGGAAAAGAAATTGCAAGGAAAATGAATTTAGCTGTGAGGTATGTGGGCAGACATTTAGAGTGCCTTT  
 TGATGTTGAGATCCACATGAGAACACAAAGATTCTTTCACTTACGGGTGTAAATGTGCGGAAGAAGATTCA  
 AGGAGCCTTGGTTTCTTAAAAATCAGATGCGGACRCATAATGGCAAATCGGGGGCCAGAAGCAAATGCAGCAA  
 GGCTTGGAGAGTAGTCCAGCAACGATCAACGAGGTGCTCCAGGTGCACGCGGCCGAGAGCATCTCCTCTCCTTG  
 CAAAATCTGCATGGTTTGTGGCTTCCTATTTCCAAATAAGAAAGTCTAATTGAGCACCGCAAGGTGCACACCA  
 AAAAACTGCTTTTCGGTACCAGCAGCGCGCAGACAGACTCTCCACAAGGAGGAATGCCGTCTCGAGGGAGGAC  
 TTCCTGCAGTTGTTCAACTTGAGACCAAATCTCACCTGAAACGGGGGAAGAAGCCTGTGATGCATCCCTCA  
 GCTCGATCCGTTACCACCTTCCAGGCTTGGCAKCTGGCTACCAAAGGAAGTTGCCATTTGCGCAAGAAGTGA  
 AGGAATTGGGGCAAGAAGGGAGCACCGACAAACGAGATTCCGAGTTCCGAGAAGGAGCTTGGAGAAAACAAATAAG  
 AACCATTTGTGACGGCCTCTCGCAAGAGAAAGAGAAGTGCAAAACCTCCACGGCGAAGCGCCCTCCGTGGACGC  
 GGATCCCAAGTTACCCAGTAGCAAGGAGAAGCCCACTCACTGCTCCGAGTGCGGCAAAGCTTTGAGAACCTACC  
 ACCAGCTGGTCTTGCACTCCAGGGTCC

SEQ. ID. No. 4

cc43 2605 bp

CAAGCTCGAAATTAACCTCACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTA  
 GTGGATCCCCGGGCTGCAGGAATTCCGACGAGCTGGGCTACTCAGATGGCGATGAGTTTCGAGTGGCCGTGG  
 CAGTATCGCTTCCACCTTCTTTACGTTACAAACCGAATGTGACACTCGGCAGAAAGCAGCTGGCCGCTGGTG  
 CTCGCTGGTCTGTCTTCTGCGCCTGCACAAACAGTCCAGCATGACGGTGATGGAAGCTCAGGAGAGCCCGC  
 TCTTCAACAACGTCAAGCTACAGCGAAAGCTTCTGTGGAGTCGATCCAGATTGTATTAGAGGAAGTGAAGAA  
 AAAGGGAACCTCGAGTGGTTGGATAAGAGCAAGTCCAGCTTCTGATCATGTGGCGGAGGCCAGAAGAATGGGG  
 GAAACTCATCTATCAGTGGGTTTCCAGGAGTGGCCAGAACAACCTCCGTCTTTACCTGTATGAACCTGACTAATG  
 GGAAGACACAGAGGATGAGGAGTTCCACGGGCTGGATGAAGCCACTCTACTGCGGGCTCTGCAGGCCCTACAG  
 CAGGAGCACAAAGCCGAGATCATCACTGTGACGATGGCCGAGGCGTCAAGTTCTTCTAGCAGGGACCTGTCTC  
 CCTTTACTTCTTACCTCCACCTTTCCAGGGCTTTCAAAAGGAGACAGACCAGTGTCCTCCCAAGACTGGATC  
 TGTGACTCCACCAGACTCAAAAGGACTCCAGTCTGAAGGCTGGGACCTGGGGATGGGTTTCTCACACCCATA  
 TGTCTGTCCCTTGGATAGGGTGAAGCTGAAGCACCAGGGAGAAAATATGTGCTTCTTCTCGCCCTACCTCCTTT  
 CCCATCCTAGACTGTCTTGGAGCCAGGGTCTGTAAACCTGACACTTTATATGTGTTTACACATGTAAGTACATA  
 CACACATGCGCCTGCAGCACATGCTTCTGTCTCTCTCCTCCACCCCTTTAGCTGCTGTGCTCCTCCTTCTC  
 AGGCTGGTGCTGGATCCTTCTAGGGGATGGGGGAAGCCCTGGCTGCAGGCAGCCTTCCAGGCAATATGAAGAT  
 AGGAGGCCACCGGCCTGGCAGTGAAGGTTGTGGCCCCACACCGATTATGATATTTAAATCTCAACTCCCAAA  
 AAAAAAAGAAAAAAGTGAAGCTAGTTCTCTCTCTCGAGAACTAGTCTCGAGTTTTTTTTTTTTTTTTTTTTT  
 TTTTTTTTTTTTTTTTTTTTTTTTTTTGGCTTTAAGGATTTATTTATTGTTTCTTCTTACAGTGTCCACTTTTCTCTAC  
 TTAATACTACTTTCCAGTCTCAGAAGCCAGAGGGAAAAAAGACCATGAATCTTCTCTCCAGATTAA  
 GTACACACTTTGGAAAAAGATTGGAAAACTTTCTGAAAAAGTTGACTGAACTCCAAACCAACATGCCATA  
 TTGTTGATGTTGCTCATGAAAAATTGTTAAAACTCTGTTCTAGATAAAGAACAGTCTCAAGTTTGTGACAGCT  
 ACACATAGTACAGGGTCCCTATGATGATTCTTCTGTAGGACGAAATAATGTAATTTTTTTCAGTTTCTGGTTT  
 ATAACCTCTCTGATCTCAGAGTTGACTGATTAAAAACCTACTCATGCAACAGAGAATAAAGCACTCATATTTT  
 TATAAATTATATGGACAAACTATTTTGGAAATCTTATCTATTGGAGACACAATATGCTGGACTAAAGCAATAA  
 TTATTTTATTTCTCAATGTCTGTGCTAACCTCAATGACTTAGAATGCTTTGCTATATTTTGCCTCTATGCCTCAA  
 CCACACTGGCTTTCTTTTAGCTTTGAACAAGCCA  
 AACTGCTTCTCTGCCTCAGGACCAGATATTTTGGGACTTCTCTTAAGAATTCTATTTCTTAACTTTTATCTGG  
 GTAACCTAGTTTTATCCAACACTTCAGATCCTGCCGTAAAACTCTTCTTATAGAAGCCTGTCTATGACACTGTC  
 TCTCTTCTCCAACATACTCACCAGCACATGTAGACTAGATTAGAACCCTCCTGTTTTTTTCTTTTCTACTTTT  
 CTCTATCATGCTTCCCTCCATTATAATATTTTATTATGTGTGTGAATGTCTGCCCAAGTCAAGTTTCTCACT  
 AAACCTATAAACTCCGTAAAGCTGGGATCCTTCCAATTTTGTATCACCCTTAGTACAGTAGGAACACAGTAAAGA  
 TTCAATTGGTATTTGTGGAATGAATGAATGAATTGTTTTGCTAGTAAAGTCTGGGGGAACCCAGGTGAGAAGAG  
 CCTAGAAAGCAGGTGGAATCCAAGGCTAGATAGACTTAGTGTTACTCAAGAAAGGCTAGCCTGAAAATAAGGT  
 TCAAAATTATAGTCAAGAAAGTCAAGACATGGGCAAGCAAGAGTGTGCTCGTGCCGAATTCGATATCAAGCT  
 TATCGATACCGTTCGACCTCGAGGGGGGGCCCCGATCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAGTG  
 GCCGTGCTTTTACAACGTCGTGACTGGGAAAAACCTGGCGTTACCCAACCTTAAT

SEQ. ID. No. 5

41.1 1288 bp

GAGGGCAGCGAGAAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGGTGCGCCCTCGCCAA  
 CCACGCCCCGGCCCTGCCATGCATCAACCACTCAGCGCCCTGCAGTCCGTCTGAAACATCACTTGGGCAAG  
 CCACGGAGCCCTTGCCTCACCTTCTGCTCCAGCCCAAGTCAAGCACAAATTTCCATGTTCCACAAGTCCGAAT  
 CTCAATGTCTAGGACAAGCCGGTCTTGTGCTCCTCCCAAGGTCAGCCAGCGTGTCCAGGCGCTACCTGTT  
 TGAGAACAGCATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAAGAAAGCCGAGTCTCGCAAGCACAACTTT  
 GTATGTCCCCACCTCAGAAGCACGCTCTGTCTGACATCGCCGACATGGTCAAAGTCTCCTCCCAAGCCACCACC  
 CCAAGCCAGCCTCTCTCTCCAGGGTCCCCCATGAAGCTGGAAATGGATGTGAGGCGCTTTGAGGATGTCTC

CAGTGAAGTCTCAACTTTTGCATAAAAGAAAAGGCCGGCAGTCCAACCTGGAATCCTCAGCATCTTCTGATTCTAC  
 AAGCCCAGTTTGCCTCGAGCCTCTTCCAGACATCAGAGGGCAAATACCTGCTGTCTGATCTGGGCCACAAAGAG  
 CGTATGCAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGCTGGCCAACGTCAAGTACCA  
 GCTTAGGAAAAACGGGCGGGACAAAATTTCTGAAAAACATGGACAAAGGCCACCCCATCTTTTATTGCAAGTACCA  
 5 GTGCCCTCCAGTTTCAAGAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGGAC  
 ATGACCCGCTTGTCTAGTGGACCAAGGTGGAGCAAGAGATCTCCCGGGTATCGTGGCTCAGAGGTC  
 TCCAGAAACAATAGCTGCCGAAGAGGACACAGACTCTAAATTCAAGTGTAAAGTTGTGCTGTGCGACATTTGTGA  
 GCAAACATGCGGTAAAACTCCACCTAAGCAAAACGCACAGCAAGTCAACCGAACACCATTACAGTTTGTAAAC  
 10 GACGTGGATGAAGAATAGCTCTGACAGGACGAATGCCTTAGTTTCCACTTTCCAGCTGGATCCCTCACACTGA  
 ACCCTTCTTCGTTGCACCATCTGCTTCTGACATTGAACCTATTGAACCTCCTCCTGACACCCTGGCTCTGAGAA  
 GACTGCCAAAAAATAAAAAAAAAAATTC

SEQ. ID. No. 6

GCAP 2820 bp

ATCCTAAGACGCACAGCCTGGGAAGCCAGCACTGGGGAAGTGGTGCTGAGGGATGTGGGTCACTGGGGTGAAGG  
 TGGAGCTTTTCAAGGTCTCCCGTCAATGCAGCTGAGTTTCTTTGGCAGGGAATTTACCAGCTGAAGAAAGCCTG  
 CCGGCGAGAGCTACAACTGAGCAAGGCCAGCTGCTCACACCCGAGGAGGTCTGGACAGGATCTTCTCCTCTGG  
 TGGATGAGAATGGAGATGGTAAGAGGGGAGAGATGGGAGAGTGTCTGCTCACTCTGCATCATCGCCACTTTCT  
 GGCCGACGTCCTTGGGCAAGGCCCTCCACCTTCCAACCTTGGGGTCTCATCTGTGAGAAGGCTGTGGAGAAG  
 20 ATGTCATGAACATAACAAAGGGACTCATGAGCACGTGTTTGTAGGAGTGAATAAAGTCTTACAGGAGTTGCTGA  
 TGGAGGCCAGGCACGCAGAATAGAAAGAATAGGAACCTTTGGAGTCAGGCAGGGAGTGATATATTGAGCTTCTCG  
 TCCTAGTCTCAATTTCTCATCTGGAAAATGGGGATAATAATAGTGGTTGAGAGGAATGAATAGGATAATGTGT  
 TTAAGAGCAGGCATAGGGTAGACCTCCATTCAAGGCTGCTTGGGCTTCTCCTCCCTGTAGCCCAAAGCCAGCCTC  
 AGGGCTATGTGGGAGAGAGCTGGCTTGGAAATACACACTTGAGCCCTCCAGCTCTCTCAGCTCCACCCAGCATT  
 25 TCCGTGGTACCATGCGCAAAAGTAAAACTTCAATTATCAGCAAGAAAGCCCTTAAAGGTGGCAGGAGACTC  
 CTGGAGATTACAGACACCTGACAAGCCGCAAGCTTGGAGTCTGAGACTGCAGGATAGTTGGCATAAGACGTGTAG  
 GCGCATCTCTGGGAGCGAGGTCTCTCTCTGCCCCAGACCCAGGTCTCCCTTCTCTACATGACACACCTCTC  
 CTCCCCCTTGTCTAGGCCAGCTGTCTCTGAACGAGTTTGTGTAAGGTGCCCGTCGGGCAAGTGGGTGATGAAG  
 ATGCTGCAGATGGACATGAATCCCAGCAGCTGGCTCGCTCAGCAGAGACGGAAGTGCATGTTCTGAGGAGT  
 30 CTGGGGCCCTCCACGACTCCAGGCTCACCCAGGTTTCCAGGGTAGTAGGAGGGTCCCTGGCTCAGCCTGCTC  
 ATGCCCACTCTTCCCTGGTGTGACTTCTTGGCACCCCTGTGTCAGGGCTGAGTGGGGATGGGGAAGGGCTGC  
 TGGGTTTGAAGTGGCCAAAGGGCATAGTCCATTTTGGAGGAGTCCCTGGGATGGTGAAGGGAATTCAGTTACT  
 TTTCTGTTTCAAGCGCTCCTGGGAGGACTGTGCTTGGCTGGGTGGTGTGGGGCTCCACAGTTTCTGGGTGT  
 TCTCAGTTGGAAGCAAGAGCCAACCTGAGGGGTGAGGGTCCACAGACCAAATCAGAAATGAGAACACAAAGACT  
 35 GGTAGGAGGCAGGGGTGGGAGGGTGTGAGACTGAAGAAAAGGCAGGAGTTGCCGGGCACGGTGGCTCACGCCT  
 GTAATCCCAGCACTTTGGGAGGCCGAGGCCGGCAGATCACAGGTCAGGAGATCGAGACCCTCTGGCTAACAC  
 GGGGTGAAAACCCGCTCTACTATAAAATACAAAAATCAGCCGGGTGAGGTGGCGGGCGCTGTAGTCCCAGCT  
 ACTCAGGAGGCTGAGGCAAGAGAATGGCGTGAACCCAGGGGGCCGAGCCTACAGTGAGCCGAGATTGCGCCAC  
 TGCACCTCCAGCCTGGACGACAGTGAGACTCCGTCTCAAAAAAAAAAAGAAAGAAAGAAAGGAGGAGTTT  
 40 TGGGGGGCAGGGGGCAGCAATAATTCTATAACTTCCGGGATGCTGAGGGGCGTTCATGGGGAGGACCTTGGCCT  
 CCTCCTCCCCAAGGCATCCTCACCAGTGGTGTCAACAGGAAAATGGCAGCAAAATAGCTGCGGCTGTGGTCT  
 TTCTGCTTTGAAAGGGTCAGCTGTACTTAAAGGGAAGTTTTCAGCTCTGCTTGGGTGCTGCTCTGGGACCCCC  
 TGCTGCCAACCCACCACTCCCCCAACAATCCTCTCTTTCCATCCATATCCCCAGTATGGACCTTCCACAACCTC  
 CCAGCCATAAGCTGAATGTTTCTCTTTAAAGGATGGAGAAAACCTTCTGTCTGTCTCTGGCAAGAAATTGGGGGAC  
 45 TGTGACTGGGATTGTGGGCTGGGCTTGGCTTCTAACTGCTGTGTGACCCAGACAGCCACTTCTCCTCCCTAA  
 CCTTGGTTATGTCTTGGCAGCACAGTGAGCAGGTCGGACTAGGCCAAGCAGTTTGGATTATGTGTTTTAGAT  
 GTGGAATTATTTTGTATATATAAACTCTTATGTGTAAACCCCAATATAGAACTAGATTAAAGGGAGTCTCTC  
 TGGTTGAAAGGGGAGCTGAGTACCTCTGGAACCTGGAGGCACCTCTGAAAAAGCAAACCTGAAAAACAGTGCCC  
 50 TGGGTCACTGTACTCTTATAAGACAGTTTAAAGTGAGACCTGGAACCAATTTGCTTTACCTTGAATAGATAG  
 GTTTTTATGTTGGTATATAAGAAATAAACTAACTATTAAACCTGAGACTTTACAGGTGTGTTATTTTCATATG  
 ATAGTCATATAAAATTTCTTTTAGACATCAATTTTAGTTAGTAAAAATAATTGATTAGAAAAATATTGGCCAGGTG  
 CAGCAGCTCACACCTGCAATCCAGGACTTTGGGAGGCCGAGGCGGGTGGATCACCTGAGGTGAGGGTTCAAG  
 ACCAGCCTG

SEQ. ID. No. 7

1b4 1205 bp

GCGCGCGTGAGTCCGCCCCCAGTCACGTGACCGCTGACTCGGGGCGTTCTCCACTATCGCTTACCTACCTCC  
 CTCTGCAGAAACCCGCGATATGGCTGCGGCTGTGCCCCGCGCGCATTTCTCTCCCCGCTGCTTCCCTTCTCC  
 60 TGGGCTTCTGCTCCTCTCCGCTCCGATGGCGGCGGCGCTGCACACCAAGGCGCCCTTCCCTGGATACGG  
 TCACCTTCTACAAGGTCATTCCCAAAGCAAGTTCGTCTGGTGAAGTTCGACACCCAGTACCCCTACGGTGAGA  
 AGCAGGATGAGTTCAAGCGTCTTCTGAAAACTCGGCTTCCAGCGATGATCTCTTGGTGGCAGAGGTGGGGATCT  
 CAGATTATGTGACAAGCTGAACATGGAGCTGAGTGAGAAATACAAGCTGGACAAAGAGAGCTACCCATCTTCTA  
 CCTCTTCCGGGATGGGGACTTTGAGAACCCAGTCCCATCACTGGGGCAGTTAGTTGGAGCCATCCAGCGCTG  
 65 GCTGAAGGGGCAAGGGGTCTACCTAGGTATGCTGCTGTATACGACGCCCTGGCCGGGGAGTTTCAT  
 CAGGGCCTCTGGTGTGGAGGCCGCCAGGCTCTTGAAGCAGGGGCAAGATAACCTCTCAAGTGTGAAGGAGAC

TCAGAAGAGTGGGCCGAGCAATACCTGAAGATCATGGGGAAGATCTTAGACCAAGGGGAGCACTTCCAGCATCA  
 GAGATGACACGGATCGCCAGGCTGATTGAGAAGAACAAGATGAGTGACGGCAGAAGGAGGAGCTCCAGAAGAGC  
 TTAAACATCCTGACTGCCCTCCAGAAAGAGGGGGCCGAGAAAGAGGAGCTGTAAAAAGGCTGTCTGTGATTTTC  
 CAGGGTTTGGTGGGGGTAGGGAGGGGANAGTTAACTGCTGGCTGTGANTCCCTTGTGGAATATAAGGGGGYMS  
 KGGGAAAAGWGGTACTAACCCACGATTCTGAGCCCTGAGTATGCCTGGACATTGATGCTAACATGACCATGCTT  
 GGGATGTCTCTAGCTGGTCTGGGGATAGCTGGAGCACTTACTCAGGTGGCTGGTGAATGACACCTCAGAAGGA  
 ATGAGTGCATAGAGAGGAGAGAGGAGTGTACTGCCAGGTCTTTGACAGATGTAATTCTCATTCAATTAAAGT  
 TTCAGTGTTTTGGTTAAGTGG

SEQ. ID. No. 8

20sa7 456 bp

GAAATCAGAAGTTTAAATATGACACAATTAAATATATTTGTATATCTCACACCGGAGNTTCTCTTCAAACATAAG  
 GAGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTAGATAAATTCATTTCGATTAAATTAAATTCAGATA  
 GAGAGAAGTAATTTTCGGAAAAGAAAATGATAGCTATATTAAGCAGATATTATTACAATACCATGTAGAGACA  
 TAAGCAATATTTTGGCATCATCTGTCCGCTCAGTAGGCCGTGTTCCCTCTGGTAGGGCCTTTGGAGAGTACCA  
 TCTATCTAAGATGGAGGAATGCTGTGGGAAGGGCGGGATGGAGGTGCGTTTTCTACGCTGAACCCACACAGGA  
 AATCTGCAGCCACACAGCTGCCTCTGCGCCGCTTCCATGTGATCATCCTGGTCAATGAAGTGAATTGTCTTA  
 TTTTCNGGGGGT

SEQ. ID. No. 9

Genomic Sequence Encoding ZABC1

CCATCATATTTCTTATTTTTTTGGGCGGAGAGGGGAGACTTGCTCTGTTGCCAGGCTGGACCAGTGGTGCGATCT  
 TGGCTCACCTGCAACCTCCACCTCCTGGGTTCAGTGATTCCCAAATAGCTGGGATTACAGGTGTGATTACCATGC  
 CCAGCTAATTTTTGTATTTTTCAGAGATAAGGGGTTTACCAGTGTGGCCAGGCTGGTCTCCAACTCCTGGCCTCA  
 TGTGATCCACCCACTTCGGCTTCCCAAAGCATTGGGAGTATAGGTGTGAGCCACTATACCCGCTCTCACATCATAT  
 TTCTAATCCCGAGACTGTAGAGCTGGTGTCTCTTTTTCTAAAGGATGTGAGTAGAGAAGTGGAGTTCCTCCAAAT  
 ACAGTTTCACGTATTAGTCAAGTTTCTAAAATACAGTAATAATGTTGAGAGCTGACATAGGGACTAACTTGGTTTT  
 TTTTTTTTTTTTTTTTTTCAAATTTCTCACTGAACTTTGATTTTGCTAAATAAGGACATTAAAAAAAACCAAAA  
 AACTCCACTATTGCCCTATTGCCACTATTTGATTTTTTAAAAAATAAGCGTATTTTAGCATCTAAAAAGTAGGAAGGA  
 CCTCAAATAAATGAGTCTTTGTTCTTGCCAGGGAAGAACAGCGTTGTCAGAATTTGATAACTGTTTTCTAGGGTA  
 TGTGCTGTTATTCAGTAAAAACCTTGCTGGGACGCTAGCATTCAGTAAATACTTGTGAATAAGCAAATGAAACT  
 TAAGCTCTATGTATAGAAACCTAAGTCACTTACATCTGATTAGCAGAGTAATTGAATATTCTTTCAATGTGT  
 AGCTCTATCCCCAGAACCACAGAATATTGGAAGTGTAAAGGCCATCCTATAGTTTAAACCACTGCGTTAAATAGAT  
 AATAGAAAGATGTGGTATGTGGCAGTGACAACCTGAAGGTGTGACTAGAAGTCTGGAGTGTCTTATTA  
 TATCACACCAAGCTGGTACCAGCCCATGTGTGATCCTCATTGTGATAGCAACAAAGAAAAGACTTCAGGACAT  
 TCTTCCCTTTACCCTAATCCTTGATCTGCAGTCTTATTTAGAAAAGCTTAATGTTAAAGATCTAGTTTATTCAAAA  
 CTAAGATAACAAGGAGTATGAGAATTTCTATTTTCGGAGTGTAAAGGAGGAGATGTTTCTTGGCTTCTCTGAGCC  
 TGCAGGCCCTTCTTGTCTTTAAGGAAGTAGAGAGAGGGAGGAAAGTAAAGTATGCTTTTGTTTTTAAGGTTACT  
 TTGCTGGGAGTAGTTTGCATGCCTTTTGGTTTTCTTGGGTGGAATTAAGTACTTAAAGTATTGTTTCTTCCAGG  
 ATGAAAATTAGCTTCTAAACCAGAAGCAGCAGGCTAAATAAAGTAATTTTCCACCTGGCCAGTGCATGATGTGA  
 AAGGTAGATTAATAAATGAGAGGGCCCATTTTCTGATGAAAGACTAAGCCATGTTGAACAGCCCTGTTGAGGAT  
 TTTATTTTAAATCTATACATTACAAAAGGAGCTTTGTGTATGTCTTTCCCTATTTGTTGTTTGGACTAGGAAGCCC  
 CACCCAGTGCTTGTGAAGGCAGAAAGTCGTTGAAAGCAAGCTGGGATTGAAACAGTGGATTGAGGTTTGAATAT  
 CCAGTGAACCAAAATATACAGGTTTCCCTGGCCAAAGATGAGTGACCATTCGAGGTGTTAAGTATTCTTGAAT  
 GGGGATTTTAGGAAAAGTTTCTGTATTTCTGTGCTCATTTTGTGACCTCTGTATGTGCAAAATCTCTAAGGGGGT  
 GTTTGGGCACTTAGATTCTTGGATGCAGATTGTTGTATATGAAACAAATTTTAAATTGTTTTGTATACACTGG  
 ATTTAAATAGTTTACTAAAGTGTTTAATTTTTTCATCTTAATTTTACAGTCTTATAGTCTTTAGATTAGGG  
 AGGCTGTTGATGGCATCCACATGTGCATTTTAGTGCCATTTAAATGTATTCAGCTGAATTTAAACAATTTCTGACC  
 TAAAACCTGACATTTTAGATTTAAGTCGGTAAAGCACTGATTAAACTGGATTTTAACTGGATGAAATTTCTGATTT  
 AATAAGTGTACTGACTGGATAAAATGCCAATGATTTAATTAACAAGCACGTTTAAACAGGATGCCCTATATATTAGT  
 TAAAAGTGAAGCAATTGAATTAGGTACCTTCTCTGCTGCGTGGAAAAGACCGTATGACTCACCCACACAGCCCTC  
 TCTTCGCTCTGAGTGTAGCTAACCGTTTCTGTTTTTCTCTAGGGTTTGGAAAATCCCTGTCTCCAGGTTGCT  
 GGGATTGCACTTCTTGCTCAATTGAAACACTTCAATGGAGACAAAGAGAACTAATGCTTTGTGCTGATTCATAT  
 TTGAATCGAGGCATTGGGAACCCGTGATGCCTTGTTTGTGGAAGAACCAGTGACACCATCACTGAGCTTCTTAAA  
 AGTTGGAAGAAGTTAGAGGACTATACACTTCTTTGAACTTTTATAATAAATATTGCTCTGGTTTTTGGAAACCC  
 AGGGCTGTAGAGGGGTGAGTGACAAGTCTTACAAGTGGCCTTATTTCAACTCCAGAAATGCCCCAACGGAATTT  
 GAGATTATATGCAATCGAAAGTGACAGGAACACTTCAATGGAGACATTTCCCTCTTAATGTACATGGATGGGCCAGAAGT  
 GATTGGCAGCTCTTGGCAGTCCGATGGAGATGGAGGATGCCTTGTCAATGAAAGGGACCGCTGTTGTTCCATTCT  
 CGAGCTACACAAGAAAAAATGTATCCAAATCGAGGGGTATATGCCCTTGGATTGCATGTTCTGCAGCCAGACCT  
 TCACACATTACAGAAGACCTTAATAAACATGTCTTAATGCAACACCGGCCCTACCCTCTGTGAACCAGCAGTTCTTCG  
 GGTGGAAGCAGAGTATCTCAGTCCGCTTGATAAAAAGTCAAGTGCGAACAGAACCTCCCAAGGAAAAGAATTGCAAG  
 GAAAATGAATTTAGTGTGAGGTATGTGGGCAGACATTTAGAGTCGCTTTTGATGTTGAGATCCACATGAGAACAC  
 ACAAAGATTCTTTCACTTACGGGTGTAACATGTGCGGAAGAAGATTCAAGGAGCCTTGGTTTCTTAAAAATCACAT  
 GCGGACACATAATGGCAAATCGGGGGCCAGAAGCAAACCTGCAGCAAGGCTTGGAGAGTAGTCCAGCAACGATCAAC



ACACAAAGATACTCAGGATTTCATATAAAATGCAGCTGATGACAGTGCTGATAAAGTGAATAAAAAACCTACCCCT  
 GCTTACCTGGACCTGTTAAAAAGAGATCAGCAGTTGAACTCAGGCAAATAACCTCATCTGTAGAACCAAGGCGG  
 ATGTTACTCCTCCTCCGGATGGCAGTACCACCCATAACCTTGAAGTTAGCCCCAAAGAGAAAGCAAACGGAGACCGC  
 AGCTGACTGCAGATACAGGCCAAGTGTGGATTGTACGAAAAACCTTTAAATTTATCCGTGGGGGCTCTTCACAAT  
 TGCCCCGCAATTTCTTTAGTAAAAAGTTTGATTCCAAGTATCACCCTGTCCATTTTGTACCTTCAAGACATTTTATC  
 CAGAAAGTTTAAATGATGCACCAGAGACTGGAGCATAAATACAATCCTGACGTTTCATAAAAACTGTGCAAAACAGTC  
 CTTGCTTAGAAGTCGACGTACCGGATGCCCGCCAGCGTTGCTGGGAAAAGATGTGCCTCCCCCTCCCTAGTTTCTGT  
 AAACCCAAAGCCCAAGTCTGCTTTCCCGGCGCAGTCCAAATCCCTGCCATCTGCGAAGGGGAAGCAGAGCCCTCCTG  
 GGCCAGGCAAGGCCCTCTGACTTCAGGGATAGACTCTAGCACTTTAGCCCCAAGTAACCTGAAGTCCCACAGACC  
 ACAGCAGAAATGTGGGGGTCCAAGGGGCCGCCACCAGGCAACAGCAATCTGAGATGTTTCTAAAACAGTGTTC  
 CCTGCACCGGATAAGACAAAAAGACCCGAGACAAAATTGAAACCTCTTCCAGTAGCTCCTTCTCAGCCACCCCTCG  
 GCAGCAGTAACATCAATGGTTCCATCGACTACCCGCCAAGAACGACAGCCCGTGGGCACCTCCGGGAAGAGACTA  
 TTTCTGTAATCGGAGTGGCAGCAATACTGCAGCAGAAATTTGGTGAGCCCCCTTCCAAAAAGACTGAAGTCCAGCGTG  
 GTTGCCTTGACGTTGACCAGCCGGGCCAATTACAGAAGAGGCTATGACCTTCCCAAGTACCATATGGTCAGAG  
 GCATCACATCACTGTTACCGCAGGACTGTGTGTATCCGTGCGAGGCGCTGCCTCCCAACCAAGGTTCTGTAGCTC  
 CAGCGAGGTGCTATTCTCAAATGTGTGACTGTTTCAAGCCCTATGGTGGCTCCGGGCCACTTTACACTTGTGTG  
 CCTGCTGGTAGTCCAGCATCCAGCTCGACGTTAGAGGTTATGATGAGGGGCGTGTGTTAAATGGCTGCCTAC  
 AGTGATTAATAGCTAATCCAGGCATTCTCAGTGGAGATGGTACCCTCCCAAGGGTGGGGGTAGGAGCCAGCAAG  
 TTCTTGGGGTTCACAGAGAGAAGCATTCTTAGATACGGCAGTGGTTTGTGGTCTTCAAGGCTTACTTAACCTGTG  
 GGGTTTAACTCTTAACCCCTGTGTATTTTATTCTTTGATTGTTTGTAGTCTTACTTTATTTTATAGAGAAAGGGTCTT  
 GCTCCGTCTCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATC  
 CTTCTGCCTCAGCTTCCAGGTAGCTGAGACTATATGTGCTGTACCATGCACAGCTGATTTTAAATTTTTTTTGT  
 TAGAGATGGAGTTGCCAGGCTGGTCTTGAACCTCTGGCTGAGGTGATCTCCTGCGTTGACCTCCCAAGTATCT  
 TAGACTACAGATGCACTCCACCACGCTTG

SEQ. ID. No. 10

ZABC1 Open reading frame

ATGCAATCGAAAGTGACAGGAAACATGCCAACTCAATCCCTCTTAATGTACATGGATGGGGCCAGAAGTGATTGGCA  
 GCTCTCTTGGCAGTCCGATGGAGATGGAGGATGCCCTGTGCAATGAAAGGGACCGCTGTTGTTCCATTCCGAGCTAC  
 ACAAGAAAAAATGTCATCCAAATCGAGGGGTATATGCCCTGGATTGCATGTTCTGCAGCCAGACCTTCACACAT  
 TCAGAAGACCTTAATAAACATGTCTTAATGCAACACCGGCCCTACCCCTCTGTGAACCAGCAGTTCTTCGGGTGAAG  
 CAGAGTATCTCAGTCCGCTTGATAAAAGTCAAGTGCGAACAGAACCTCCCAAGGAAAAGAATTGCAAGGAAAATGA  
 ATTTAGCTGTGAGGTATGTGGGCAGACATTTAGAGTCGCTTTTGTATGTGAGATCCACATGAGAACACACAAAGAT  
 TCTTTCACTTACGGGTGTAACATGTGCGGAAGAAGMTTSSRSSAGCCTTGGTTTCTAAAAATCACATGCGGACAC  
 ATAATGGCAAAATCGGGGGCCAGAAGCAAACTGCAGCAAGCCTTGGAGAGTAGTCCAGCAACGATCAACGCGTGT  
 CCAGGTGCACGCGGCCGAGAGCATCTCTCTCTTACAAAATCTGCATGGTTTGTGGCTTCTTATTTCCAAATAAA  
 GAAAGTCTAATTGAGCACCGCAAGGTGCACACCAAAAAAACTGCTTTCGGTACCAGCAGCGCGCAGACAGACTCTC  
 CACAAGGAGGAATGCCGTCTCGAGGGAGGACTTCTGCACTTGTTCAACTTGAGACCAAAATCTCACCTGAAAC  
 GGGGAAGAAGCCTGTCAGATGCATCCCTCAGCTCGATCCGTTTACCACCTTCCAGGCTTGGCAGCTGGCTACCAAA  
 GGAAAAGTTGCCATTTGCCAAGAAAGTGAAGGAATCGGGGCAAGAAGGGAGCACCGACAACGAGATTGAGTTCCG  
 AGAAGGAGCTTGGAGAAACAAATAAGGGCAGTTGTGCAGGCCTCTCGCAAGAGAAAGAGAAGTGCAAACTCCCA  
 CGGCGAAGCGCCCTCCGTGGACGCGGATCCCAAGTTACCCAGTAGCAAGGAGAAGCCCACTACTGTCCGAGTGC  
 GGCAAGCTTTTCAAGCTTACCACAGCTGGTCTTGCACTCCAGGCTCCACAAGAAGGACCGGAGGCGGGCGG  
 AGTCGCCACCATTCTGTGGACGGGAGGCGCGGGAGCTGTTTCTCTGACCTCGCCGCCCTCTGGATGAAAA  
 TGGAGCCGTGGATCGAGGGGAAGGTGGTTCTGAAGACGGATCTGAGGATGGGCTTCCCGAAGGAATCCATCTGGAT  
 AAAAATGATGATGGAGGAAAAATAAAACATCTTACATCTTCAAGAGAGTGTAGTTATTGTGGAAGTTTTCGGTT  
 CAAATTATTACCTCAATATTCTCTCAGAACGCATACAGGTGAAAAACCATACAAATGTGAATTTTGTGAATATGC  
 TGCAGCCCAAGACATCTCTGAGGTATCACTTGGAGAGACATCACAAGGAAAAACAAACGATGTTGCTGTGAA  
 GTCAAGAAACGATGGTAAAAATCAGGACACTGAAGATGCATTAACCGCTGACAGTGCAGCAAAACCAAAATTTGA  
 AAAGATTTTTTGTATGGTGCCAAAGATGTTACAGGCAGTCCACCTGCAAGCAGCTTAAGGAGATGCCTTCTGTTTT  
 TCAGAATGTTCTGGGCAGCGCTGTCTCTCACCAGCACAAAGATACTCAGGATTTCCATAAAAAATGCAGCTGAT  
 GACAGTGCTGATAAAGTGAATAAAAAACCTACCCCTGCTTACCTGGACCTGTTAAAAAAGAGATCAGCAGTTGAAA  
 CTCAGGCAAAATAACCTCATCTGTAGAACCAAGGCGGATGTTACTCTCTCCGGATGGCAGTACCACCCATAACCT  
 TGAAGTTAGCCCCAAAGAGAAGCAACGAGACCGCAGCTGACTGCAGATACAGGCCAAGTGTGGATTGTCACGAA  
 AAACCTTTAAATTTATCCGTGGGGGCTCTTCACAATTGCCCGCAATTTCTTTGAGTAAAAAGTTTGATTCCAAGTA  
 TCACCTGTCCATTTGTACCTTCAAGACATTTTATCCAGAAAGTTTAAATGATGCACCAGAGACTGGAGCATAAATA  
 CAATCCTGACGTTTCAAAAACTGTGAAACAAGTCTTGTGTAAGTTCGACGTACCGGATGCCCGCCGAGCGGTTG  
 CTGGGAAAAGATGTGCTCTCTCTAGTTTCTGTAAACCAAGCCCAAGTCTGCTTTCCCGGCGCAGTCCAAAT  
 CCCTGCCATCTGCGAAGGGGAAGCAGAGCCCTCTGGGCCAGGCAAGGCCCTCTGACTTCAGGGATAGACTCTAG  
 CACTTTAGCCCCAAGTAACCTGAAGTCCCACAGACCAAGCAGAAATGTGGGGGTCCAAGGGGCCGCCACCAGGCAA  
 CAGCAATCTGAGATGTTTCTAAAACAGTGTTCCTGTCACCGGATAAGACAAAAAGACCCGAGACCAAAATGA  
 AACCTCTTCCAGTCTCTTCTCAGCCCCACCTCGGCAGCAGTAACATCAATGGTTCCATCGACTACCCCGCAA  
 GAACGACAGCCCGTGGGCACCTCCGGGAAGAGACTATTTCTGTAATCGGAGTGCCAGCAATACTGCAGCAGAAAT  
 GGTGAGCCCTTCCAAAAAGACTGAAGTCCAGCGTGGTTGCCCTTGACGTTGACAGCCCGGGGCCAATTACAGAA  
 GAGGCTATGACCTTCCCAAGTACCATATGGTCAGAGGCATCAGTCTGTTACCGCAGGACTGTGTGTATCCGTC  
 GCAGGCGTGCCTCCCAACCAAGGTTCTCAGCTCCAGCAGGTCGATTTCTCAAATGTGCTGACTGTTTCAAGG  
 CCTATGCTGGCTCCGGGCCACTTTACACTTGTGTGCTGCTGGTAGTCCAGCATCCAGCTCGACGTTAGAAGGTC  
 TTGGTGGATGTCAGTGCTTACTCCCCATGAAATTAATTTTACTTCATCCTTTGAGAAGCGAATGGTGAAAGCTAC



TGAAATAAGCTGTGATTGTACTGTACATAAAACATATGAGGAATCTGCAAGGAACACTACAGTTGTGTAA

SEQ. ID. No. 11

ZABC1 Protein

5 MQSKVTGNMPTQSLLMYMDGPEVIGSSSLGSPMEMEDALSMKGTAVVPFRATQEKVNIQIEGYMPLDCMFCSTFTTH  
 SEDLNKHLVLMQHRPTLCEPAVLRVAEYLSPLDKSQVRTEPPKEKNCKENEFSCVCGQTFRVAFDVEIHMRTHKD  
 SFTYGCNMCGRXXXXPWFLKNHMRTHNGKSGARSKLQQGLESPATINEVVQVHAAESSIPYKICMVCGLFPNK  
 10 ESLIEHRKVHTKKTAFGTSSAQTDSPQGGMPSSREDFLQLFNLRPKSHPETGKKPVRCIPQLDPFTTFQAWQLATK  
 GKVAICQEVKESGQEGSTDNDSSSEKELGETNKGSCAGLSQEKEKCKHSHGEAPSVADADPKLPSSKEKPTHSEC  
 GKAFRTYHQLVLHSRVHKKDRRAGAESPTMSVDGRQPGTCSPLAAPLDENGAVDRGEGGSSEDGSEDGLPEGIHL  
 KNDDGGKIKHLTSSRECSYCGKFFRSNYLNIHLRTHTEKPKYKECFEYAAAQKTSRLYHLERHHKEKQTDVAAE  
 VKNDGKNQDTEADALLTADSAQTKNLKRFFDGAKDVTGSPPAKQLKEMPVVFQNVLGSVLSPAHKDQDFHKNAAAD  
 15 DSADKVNKNPTPAYLDLLKKRSVETQANNLICRTKADVTPPPDGSTTHNLEVPKKEQTETAADCRYRPSVDCHE  
 KPLNLSVGALHNCPAISLSKSLIPSITCFCTFKTFYPEVLMMHQRLEHKYNPDVHKNCNRNKSLLSRRTGCPPAL  
 LGKDVPPLSSFCPKPKPSAFPAQSKSLPSAKGKQSPGPGKAPLTSIGDSSTLAPSNLKSHPQQNVGVQGAATRO  
 QQSEMFPKTSVSPAPDKTKRPETKLPLPVAPSQPTLGSSNINGSIDYPAKNDSPWAPPGRDYFNRSASNTAAEF  
 GEPLPKRLKSSVVALDVDPGANYYRRGYDLPKYHVMVRGITSLLPQDCVYPSQALPPKPRFLSSSEVDSPNVLTVQK  
 PYGGSGPLYTCVPAGSPASSSTLEGLGGCQCLLPMKLNFTSSFEKRMVKATEISCDCTVHKTYEESARNTTVV

20 SEQ. ID. NO. 12

1b1

GGAAACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCCTCACTAAAGGGAACAAAAGCTGGAGCTCCACC  
 GCGGTGGCGGCGGCTCTAGAACTAGTGATCCCCGGGCTGCAGGAATTCGGCACGAGGCTCCACCGACAGCCAGG  
 25 CACTGGGCAGCAGCACTGGAGACCCAGGACCCTGTGCAGGAGCAGCTCCGGGTGACACGAGGGGACTGAAGATAC  
 TCCCACAGGGGCTCAGCAGGAGCAATGGGTAACCAATGAGTGTTCCCCAAAGAGTTGAAGACCAAGAGAATGAAC  
 CAGAAGCAGAGACTTACCAGGACAACGCGTCTGCTCGAACGGGGTTCCAGTGGTGGTGTGCGACCCACACAGTTCA  
 GCACTTAGAGGAAGTCGACTTGGGAATAAGTGTCAAGACGGATAATGTGGCCACTTCTTCCCCGAGACAACGGAG  
 ATAAGTGCTGTTCGGATGCCAACGGAAAGAATCTTGGGAAAGAGGGCCAAACCCGAGGCACCAGCTGCTAAATCTC  
 GTTTTTCTTGATGCTCTCTCGGCTGTACCAGGACGTACCGGAGACCAAGCCGAGATTCTATCCCTTGGATCAGT  
 30 GAAGCTTGATGTCACTCCAATAAAGCTCCAGCGAACCAAGACCCAAAGTGAGAGCTGGACACTTCCGGTGGCAGCT  
 GGACCTGGGCAGGACACAGATAAAACCCAGGGCAGCCCCGGCCCAAGACAAGGTCTCTCTGCCGCCAGGGATC  
 CCACGCTTCTCCACCTGAGACAGGGGGAGCAGGAGGAGAAGCTCCCTCCAAGCCCAAGGACTCCAGCTTTTTTGA  
 CAAATCTTCAAGCTGGACAAGGGACAGGAAAAGGTGCCAGGTGACAGCCAACAGGAAGCCAAGAGGGCAGAGCAT  
 35 CAAGACAAGGTGGATGAGGTCTCGCTTATCAGGGCAGTCCGATGATGTCCCTGCAGGGAAGGACATAGTTGACG  
 GCAAGGAAAAAAGGACAAGAACTTGGAACTGCGGATTGCTCTGTCCCTGGGGACCCAGAAGGACTGGAGACTGC  
 AAAGGACGATTCCCAGGCAGCAGCTATAGCAGAGAATAATAATTCCATCATGAGTTCTTTAAACCTCTGGTTTCA  
 CCTAACAAAGCTGAAACAAAAAGGACCCAGAAGACACGGGTGTGAAAAAGTCAACCCACCTTACAGTGACCTTA  
 AGTCAGACAAAGCCAACCTTACATCCAGGAGACCCAAGGGGCTGGCAAGAATTCCAAAGGATGCAACCCATCGGG  
 40 GCACACACAGTCCGTGACAAACCCCTGAACCTGCGAAGGAAGGCACCAAGGAGAAATCAGGACCCACCTCTCTGCCT  
 CTGGGCAAACTGTTTTGAAAAAGTCAGTTAAAGAGGACTCAGTCCCCACAGGTGCGGAGGAGAAATGTGGTGTGTG  
 AGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTACAAACAGTGGACCTCAACGAAGGAGATGC  
 TGCACCTGAACCCACAGAAGCGAAACTCAAAAGAGAAGAAAGCAAAACCAAGAACCTCTCTGATGGCGTTTCTCAGA  
 CAAATGTCAGTGAAAGGGGATGGAGGATCACCCACTCAGAAGAAATAAATGGGAAAGACTCCAGCTGCCAAACAT  
 45 CAGACTCCACAGAAAAGACTATCACACCGCCAGAGCCTGAACCAACAGGAGCACCACAGAAGGGTAAAGAGGGCTC  
 CTCGAAGGACAAGAAGTCAGCAGCCGAGATGAACAAGCAGAAGAGCAACAAGCAGGAAGCCAAAAGAACAGCCAG  
 TGCACAGAGCAGGCCACGGTGGACACGAACCTCACTGCAGAATGGGGACAAGCTCCAAAAGAGACCTGAGAAGCGGC  
 AGCAGTCCCTTGGGGGCTTCTTTAAAGGCCTGGGACCAAAAGCGGATGTGGATGCTCAAGTGCAAAACAGACCCAGT  
 50 ATCCATCGGACCAAGTTGGCAAAACCAAGTAAACAAATCAGCACGGTTCACACAGGTTCTCCTGCCACCAAGATGT  
 GTTCTCCTTACTCCATCTCTCTCCCAACACGCTCCATGTATATATTCTTCTGATGGCCAGCAAATGAAATTTCTGC  
 CTAGAAATTAAGCCCGAGCTGTTGTATATTGAGGTGTATTATTTACGTCTCTGGTCCAGTCTTTTCTGGCAATAA  
 CAGTAAAGATGGTTTAGCAGGTACCTAGTTGGGTGAGAAGAGTCGATGATACCAAGCAGGAAAGGGAGGGAATA  
 GAGGAATGTGTTCCGGTTAAGTGATGAAAATGGCAGTGGTGGCCGGGCGTGGTGGCTCTCGCCTGTAATCTCAGCA  
 CTTTGGGAGGCCGAGGCAGGTGGATCACCTGAGGTGAGGAGTTCAAGACTAGCCTGGCCAACATCATGAAACCCCG  
 55 TCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCACACACCTGTAGTCCAGCTACTCGGGAGCCCAACGC  
 ACGAGAACCGCTGTACCCAGGAGGTGGAGGTTGCAGTGAGCCGAAGTTGCACCATTCACCTCCACCTGGGCGAC  
 AGAGCAAGATTCTATCAAAAAAAGGAGTGGCAAGTAAGTTATAGAAGAGAAATGCTGCTAGAAAGGAATTAAG  
 CGTTGTGATAACGCGTGTCTCTCTAAGCTTGAAGAGGAGACGAAAATCCATTTGTTTAAATTCACATCTC  
 AAGGAGGGGAGAACCCGGGCTGTGTTGGGTGGTTGCCAATTTCTAGAACGGAATGTGTGGGGTATAGAAAAAGGAA  
 60 TGAATAAGCGTTGTTTTCAAAATAGGGTCCTGTGAAGTTATTGATGAGAGGGAAGATTGACTGGGGAGGGCTTA  
 AAATGATTTGGGAAAACAATTGCTTTTGAGGCTCAGTGACAACGGCAAGATTACAACCTAAAAAAGGAAATTAAG  
 AAACTCGAGACTAGTTCTCTCTCTCTCTCGTGGCAATTTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGGG  
 GGGCCCGGTACCCAATTGCGCCTATA

65



Year	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	

5 TGTGATATTGATTTCATGCCCTCTTGCACCTTGCCAAACATCACACGCTTG  
CCATCCAGTCCACTCGATTTTGGCAGTGCAGATGAAAACTGGGAACCAT  
TTGTGTTGAGTCCAGCAAGATGCCAGGACCTGCATGTTTCAGAACGAAGT  
TCTTCATCATCCAATTTCTCCCTGTATATGGGCTTACCACNACTGCCGTT  
AAGTCGTGTNAAGTCACCACTCAGGTACATAATGGAATAATTCTGCAAAG  
10 GCAGGAGNCACTTTCTCTCCAGTGCTCAGACCATGAAAGTTTTCTGATGT  
CTTTGGAACTTTGTCTGCAAATAGCTCGAAGGAGACATGGCCTAAAGGCT  
CGCCATCTGCGGTGATATTGNAACATGGTAGGGCTGACCGTGGCTGTGGC  
CATGACTTTTTAGANTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
15 NNNNNNNNNNNNNNNNNCCCAATGCGGGACAGAGAATCNAAGAACTGTATTA  
GGGAAAGGGTCCTGAGTTTATGCCAAAGTTTCCCAGATTGGT TTCCATTG  
AAACGTAGCTCTGTGAGATACCATCAGGTGTTATGTGAAGAAATGTCTGT  
GTAGTCAAATATGTTTGAGTGAGTGAGCCTGAGCTGAGCAAGACTTTACT  
GCAAGACTTCCCATCTTCTGTCCCTTTTTATGCTAATGGGTAACACAAAC  
20 TCCAAAAGTGGGGTGTACAGCATGAGGCATTAACAAAATTTATTGGACC  
CCACACACNN  
NN  
NN

gb|M19533|RATCYCA Rat cyclophilin mRNA, complete cds.  
Length = 743

Score = 418 (115.5 bits), Expect = 1.5e-58, Sum P(5) = 1.5e-58  
Identities = 96/112 (85%), Positives = 96/112 (85%), Strand = Minus / Plus  
S = rat CYCLOPHILLIN; q= SEQ ID NO 13.

```

Query:      312 AAAGTTCCAAAGACATCAGAAAACCTTCATGGTCTGAGCACTGGAGAGAAAG 261
              |||
Sbjct:     124 AAAGTTCCAAAGACAGCAGAAAACCTTCGTGCTCTGAGCACTGGGGAGAAAG 175

```

```

Query:      117 TGCTGGACTCAACACAAATGGTTCCAGTTTTTTCATCTGCACTGCCAAAATCGAGTGG 60
            |||
Sbjct:     348 TGCTGGACCAAACACAAATGGTTCCAGTTTTTTCATCTGCACTGCCAAGACTGAGTGG 405

```

```
Query:      60 GACTGGATGGCAAGCGTGTGATGTTTGGCAAGGTGCAAGAGGGCATGA 13
             | | | | | | | | | | | | | | | | | | | |
Sbjct:    404 GGCTGGATGGCAAGCATGTGGTCTTTGGGAAGGTGAAAGAAGGCATGA 451
```

Score = 154 (42.6 bits), Expect = 1.5e-58, Sum P(5) = 1.5e-58  
Identities = 34/38 (89%), Positives = 34/38 (89%), Strand = Minus / Plus

5 Query: 153 AGAACTTCGTTCTGAAACATGCAGGTCCTGGCATCTTG 116  
Sbjct: 299 AGAACTTCATCCTGAAGCATAACAGGTCCTGGCATCTTG 336

Score = 86 (23.8 bits), Expect = 1.5e-58, Sum P(5) = 1.5e-58  
Identities = 22/28 (78%), Positives = 22/28 (78%), Strand = Minus / Plus

10 Query: 256 TCCTGCCTTTGCAGAATTATTCATTAT 229  
Sbjct: 193 TCCTCCTTTCACAGAATTATTCAGGAT 220

15

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295  
1296  
1297  
1298  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564  
1565  
1566  
1567  
1568  
1569  
1570  
1571  
1572  
1573  
1574  
1575  
1576  
1577  
1578  
1579  
1580  
1581  
1582  
1583  
1584  
1585  
1586  
1587  
1588  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1598  
1599  
1600  
1601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1610  
1611  
1612  
1613  
1614  
1615  
1616  
1617  
1618  
1619  
1620  
1621  
1622  
1623  
1624  
1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638  
1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
1652  
1653  
1654  
1655  
1656  
1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680  
1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702  
1703  
1704  
1705  
1706  
1707  
1708  
1709  
1710  
1711  
1712  
1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727  
1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748  
1749  
1750  
1751  
1752  
1753  
1754  
1755  
1756  
1757  
1758  
1759  
1760  
1761  
1762  
1763  
1764  
1765  
1766  
1767  
1768  
1769  
1770  
1771  
1772  
1773  
1774  
1775  
1776  
1777  
1778  
1779  
1780  
1781  
1782  
1783  
1784  
1785  
1786  
1787  
1788  
1789  
1790  
1791  
1792  
1793  
1794  
1795  
1796  
1797  
1798  
1799  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820  
1821  
1822  
1823  
1824  
1825  
1826  
1827  
1828  
1829  
1830  
1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838  
1839  
1840  
1841  
1842  
1843  
1844  
1845  
1846  
1847  
1848  
1849  
1850  
1851  
1852  
1853  
1854  
1855  
1856  
1857  
1858  
1859  
1860  
1861  
1862  
1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900  
1901  
1902  
1903  
1904  
1905  
1906  
1907  
1908  
1909  
1910  
1911  
1912  
1913  
1914  
1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922  
1923  
1924  
1925  
1926  
1927  
1928  
1929  
1930  
1931  
1932  
1933  
1934  
1935  
1936  
1937  
1938  
1939  
1940  
1941  
1942  
1943  
1944  
1945  
1946  
1947  
1948  
1949  
1950  
1951  
1952  
1953  
1954  
1955  
1956  
1957  
1958  
1959  
1960  
1961  
1962  
1963  
1964  
1965  
1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981  
1982  
1983  
1984  
1985  
1986  
1987  
1988  
1989  
1990  
1991  
1992  
1993  
1994  
1995  
1996  
1997  
1998  
1999  
2000  
2001  
2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028  
2029  
2030  
2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089  
2090  
2091  
2092  
2093  
2094  
2095  
2096  
2097  
2098  
2099  
2100  
2101  
2102  
2103  
2104  
2105  
2106  
2107  
2108  
2109  
2110  
2111  
2112  
2113  
2114  
2115  
2116  
2117  
2118  
2119  
2120  
2121  
2122  
2123  
2124  
2125  
2126  
2127  
2128  
2129  
2130  
2131  
2132  
2133  
2134  
2135  
2136  
2137  
2138  
2139  
2140  
2141  
2142  
2143  
2144  
2145  
2146  
2147  
2148  
2149  
2150  
2151  
2152  
2153  
2154  
2155  
2156  
2157  
2158  
2159  
2160  
2161  
2162  
2163  
2164  
2165  
2166  
2167  
2168  
216

**WHAT IS CLAIMED IS:**

1                   1.     An isolated nucleic acid molecule comprising a  
 2 polynucleotide sequence having a subsequence which specifically hybridizes  
 3 under stringent conditions to a sequence selected from the group consisting of  
 4 SEQ. ID. No. 2, SEQ. ID. No. 3, SEQ. ID. No. 4, SEQ. ID. No. 5, SEQ.  
 5 ID. No. 6, SEQ. ID. No. 7, SEQ. ID. No. 8, SEQ. ID. No. 9, SEQ. ID. No.  
 6 10, SEQ. ID. No. 12, AND SEQ. ID. No. 13.

1                   2.     The isolated nucleic acid of claim 1, wherein the  
 2 subsequence specifically hybridizes under stringent conditions to SEQ. ID. No.  
 3 2.

1                   3.     The isolated nucleic acid of claim 2, wherein the  
 2 subsequence is SEQ. ID. No. 2.

1                   4.     The isolated nucleic acid of claim 1, wherein the  
 2 subsequence specifically hybridizes to SEQ. ID. No. 3.

1                   5.     The isolated nucleic acid of claim 4, wherein the  
 2 polynucleotide is SEQ. ID. No. 3.

1                   6.     The isolated nucleic acid of claim 1, wherein the  
 2 subsequence specifically hybridizes under stringent conditions to SEQ. ID. No.  
 3 4.

1                   7.     The isolated nucleic acid of claim 6, wherein the  
 2 subsequence is SEQ. ID. No. 4.

1                   8.     The isolated nucleic acid of claim 1, wherein the  
 2 subsequence specifically hybridizes under stringent conditions to SEQ. ID. No.  
 3 5.

1                   9.     The isolated nucleic acid of claim 8, wherein the  
2     subsequence is SEQ. ID. No. 5.

1                   10.    The isolated nucleic acid of claim 1, wherein the  
2     subsequence specifically hybridizes under stringent conditions to SEQ. ID. No.  
3     6.

1                   11.    The isolated nucleic acid of claim 10, wherein the  
2     subsequence is SEQ. ID. No. 6.

1                   12.    The isolated nucleic acid of claim 1, wherein the  
2     subsequence specifically hybridizes under stringent conditions to SEQ. ID. No.  
3     7.

1                   13.    The isolated nucleic acid of claim 12, wherein the  
2     subsequence is SEQ. ID. No. 7.

1                   14.    The isolated nucleic acid of claim 1, wherein the  
2     subsequence specifically hybridizes under stringent conditions to SEQ. ID. No.  
3     8.

1                   15.    The isolated nucleic acid of claim 14, 16, 18, 20, wherein  
2     the subsequence is SEQ. ID. No. 8.

1                   16.    The isolated nucleic acid of claim 1, wherein the  
2     subsequence specifically hybridizes under stringent conditions to SEQ. ID. No.  
3     9.

1                   17.    The isolated nucleic acid of claim 16, wherein the  
2     subsequence is SEQ. ID. No. 9.

1                   18.    The isolated nucleic acid of claim 1, wherein the  
2   subsequence specifically hybridizes under stringent conditions to SEQ. ID. No.  
3   10.

1                   19.    The isolated nucleic acid of claim 18, wherein the  
2   subsequence is SEQ. ID. No. 10.

1                   20.    The isolated nucleic acid of claim 1, wherein the  
2   subsequence specifically hybridizes under stringent conditions to SEQ. ID. No.  
3   12.

1                   21.    The isolated nucleic acid of claim 20, wherein the  
2   subsequence is SEQ. ID. No. 12.

1                   22.    The isolated nucleic acid of claim 1, wherein the  
2   subsequence specifically hybridizes under stringent conditions to SEQ. ID. No.  
3   13.

1                   23.    The isolated nucleic acid of claim 22, wherein the  
2   subsequence is SEQ. ID. No. 12.

1                   24.    The isolated nucleic acid of claim 1, further comprising a  
2   promoter sequence operably linked to the polynucleotide sequence.

1                   25.    The isolated nucleic acid of claim 1, which nucleic acid is  
2   a cDNA molecule.

1                   26.    A method of screening for neoplastic cells in a sample, the  
2 method comprising:

3                   contacting a nucleic acid sample from a human patient with a  
4 probe which hybridizes selectively to a target polynucleotide sequence  
5 comprising a sequence selected from the group consisting of SEQ. ID. No. 1,  
6 SEQ. ID. No. 2, SEQ. ID. No. 3, SEQ. ID. No. 4, SEQ. ID. No. 5, SEQ.  
7 ID. No. 6, SEQ. ID. No. 7, SEQ. ID. No. 8, SEQ. ID. No. 9, SEQ. ID. No.  
8 10, SEQ. ID. No. 11, SEQ. ID. No. 12, and, SEQ. ID. No. 13 wherein the  
9 probe is contacted with the sample under conditions in which the probe  
10 hybridizes selectively with the target polynucleotide sequence to form a stable  
11 hybridization complex; and

12                   detecting the formation of a hybridization complex.

1                   27.    The method of claim 26, wherein the nucleic acid sample  
2 is from a patient with breast cancer.

1                   28.    The method of claim 26, wherein the nucleic acid sample  
2 is a metaphase spread or a interphase nucleus.

1                   29.    The method of claim 26, wherein the probe comprises a  
2 polynucleotide sequence as set forth in SEQ. ID. No. 1.

1                   30.    The method of claim 26, wherein the probe comprises a  
2 polynucleotide sequence as set forth in SEQ. ID. No. 2.

1                   31.    The method of claim 26, wherein the probe comprises a  
2 polynucleotide sequence as set forth in SEQ. ID. No. 3.

1                   32.    The method of claim 26, wherein the probe comprises a  
2 polynucleotide sequence as set forth in SEQ. ID. No. 4.

1                   33.    The method of claim 26, wherein the probe comprises a  
2 polynucleotide sequence as set forth in SEQ. ID. No. 5.

1                   34.    The method of claim 26, wherein the probe comprises a  
2 polynucleotide sequence as set forth in SEQ. ID. No. 6.

1                   35.    The method of claim 26, wherein the probe comprises a  
2 polynucleotide sequence as set forth in SEQ. ID. No. 7.

1                   36.    The method of claim 26, wherein the probe comprises a  
2 polynucleotide sequence as set forth in SEQ. ID. No. 8.

1                   37.    The method of claim 26, wherein the probe comprises a  
2 polynucleotide sequence as set forth in SEQ. ID. No. 9.

1                   38.    The method of claim 26, wherein the probe comprises a  
2 polynucleotide sequence as set forth in SEQ. ID. No. 10.

1                   39.    The method of claim 26, wherein the probe comprises a  
2 polynucleotide sequence as set forth in SEQ. ID. No. 12.

1                   40.    The method of claim 26, wherein the probe comprises a  
2 polynucleotide sequence as set forth in SEQ. ID. No. 13.

1                   41.    The method of claim 26, wherein the probe is used to  
2 identify the presence of a mutation in the target polynucleotide sequence.

1                   42.    A method for detecting a neoplastic cell in a biological  
2 sample, the method comprising:  
3                   contacting the sample with an antibody that specifically binds a  
4 polypeptide antigen encoded by a polynucleotide sequence comprising a  
5 sequence selected from the group consisting of SEQ. ID. No. 1, SEQ. ID. No.  
6 2, SEQ. ID. No. 3, SEQ. ID. No. 4, SEQ. ID. No. 5, SEQ. ID. No. 6, SEQ.  
7 ID. No. 7, SEQ. ID. No. 8, SEQ. ID. No. 9, SEQ. ID. No. 10, SEQ. ID. No.  
8 12, and SEQ. ID. No. 13; and  
9                   detecting the formation of an antigen-antibody complex.

1                   43.    The method of claim 42, wherein the sample is from  
2 breast tissue.

1                   44.    A method of inhibiting the pathological proliferation of  
2 cancer cells, the method comprising inhibiting the activity of a gene product of  
3 an endogenous gene having a subsequence which hybridizes under stringent  
4 conditions to a sequence selected from the group consisting of SEQ. ID. 1,  
5 SEQ. ID. No. 2, SEQ. ID. No. 3, SEQ. ID. No. 4, SEQ. ID. No. 5, SEQ. ID.  
6 No. 6, SEQ. ID. No. 7, SEQ. ID. No. 8, SEQ. ID. NO. 9, SEQ. ID. NO. 10,  
7 SEQ. ID. No. 12, and SEQ. ID. No. 13.

1                   42.    A method of detecting a cancer, said method comprising  
2 detecting the overexpression of a protein encoded in a 20q13 amplicon.

1                   43.    The method of claim 41, wherein said protein encoded in  
2 a 20q13 amplicon is ZABC1.

1                   44.    The method of claim 41, wherein said protein encoded in  
2 a 20q13 amplicon is 1b1.



**GENES FROM THE 20q13 AMPLICON AND THEIR USES****ABSTRACT OF THE DISCLOSURE**

5 The present invention relates to cDNA sequences from a region of amplification on chromosome 20 associated with disease. The sequences can be used in hybridization methods for the identification of chromosomal abnormalities associated with various diseases. The sequences can also be used for treatment of diseases.

K:\DATA\PS\PAT\KLB\63444.1

## DECLARATION

As a below named inventor, I declare that:

My residence, post office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **GENES FROM THE 20q13 AMPLICON AND THEIR USES** the specification of which     is attached hereto or   x   was filed on 1/17/97 as Application No. 08/785,532 and was amended on     (if applicable).

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56. I claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

## Prior Foreign Application(s)

Country	Application No.	Date of Filing	Priority Claimed Under 35 USC 119
			Yes <u>   </u> No <u>   </u>
			Yes <u>   </u> No <u>   </u>

I hereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below:

Application No.	Filing Date

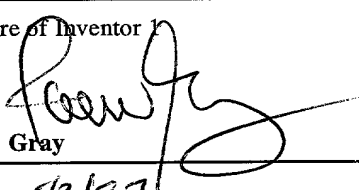
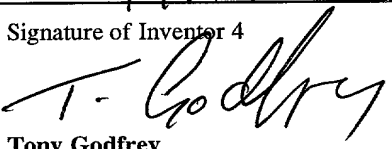
I claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, section 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application No.	Date of Filing	Status
<b>08/731,499</b>	<b>10/16/96</b>	<u>   </u> Patented <u>  x  </u> Pending <u>   </u> Abandoned
<b>08/680,395</b>	<b>7/15/96</b>	<u>   </u> Patented <u>  x  </u> Pending <u>   </u> Abandoned

Full Name of Inventor 1	Last Name <b>Gray</b>	First Name <b>Joe</b>	Middle Name or Initial <b>W.</b>	
Residence & Citizenship	City <b>San Francisco</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>USA</b>	
Post Office Address	Post Office Address <b>1921 11th Avenue</b>	City <b>San Francisco</b>	State/Country <b>California</b>	Zip Code <b>94116</b>
Full Name of Inventor 2	Last Name <b>Collins</b>	First Name <b>Colin</b>	Middle Name or Initial <b>Conrad</b>	
Residence & Citizenship	City <b>San Rafael</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>USA</b>	
Post Office Address	Post Office Address <b>333 Mountain View Avenue</b>	City <b>San Rafael</b>	State/Country <b>California</b>	Zip Code <b>94901</b>

Full Name of Inventor 3	Last Name <b>Hwang</b>	First Name <b>Soo-in</b>	Middle Name or Initial	
Residence & Citizenship	City <b>Berkeley</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>USA</b>	
Post Office Address	Post Office Address <b>189 Fairlawn Dr.</b>	City <b>Berkeley</b>	State/Country <b>California</b>	Zip Code <b>94708</b>
Full Name of Inventor 4	Last Name <b>Godfrey</b>	First Name <b>Tony</b>	Middle Name or Initial	
Residence & Citizenship	City <b>San Francisco</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>UK</b>	
Post Office Address	Post Office Address <b>699 Teresita Boulevard</b>	City <b>San Francisco</b>	State/Country <b>California</b>	Zip Code <b>94127</b>
Full Name of Inventor 5	Last Name <b>Kowbel</b>	First Name <b>David</b>	Middle Name or Initial	
Residence & Citizenship	City <b>Oakland</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>USA</b>	
Post Office Address	Post Office Address <b>6009 Auburn Avenue</b>	City <b>Oakland</b>	State/Country <b>California</b>	Zip Code <b>94618</b>
Full Name of Inventor 6	Last Name <b>Rommens</b>	First Name <b>Johanna</b>	Middle Name or Initial	
Residence & Citizenship	City <b>Toronto</b>	State/Foreign Country <b>Ontario, Canada</b>	Country of Citizenship <b>Canada</b>	
Post Office Address	Post Office Address <b>717 Bay St. #1501</b>	City <b>Toronto</b>	State/Country <b>Ontario, Canada</b>	Zip Code <b>M5G 2J9</b>

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 1  <b>Joe W. Gray</b>	Signature of Inventor 2  <b>Colin Conrad Collins</b>	Signature of Inventor 3  <b>Soo-in Hwang</b>
Date: <b>5/2/97</b>	Date:	Date:
Signature of Inventor 4  <b>Tony Godfrey</b>	Signature of Inventor 5  <b>David Kowbel</b>	Signature of Inventor 6  <b>Johanna Rommens</b>
Date: <b>5/2/97</b>	Date:	Date:



Attorney Docket No. 023070-068920

**DECLARATION**

I, a below named inventor, I declare that:

My residence, post office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **GENES FROM THE 20q13 AMPLICON AND THEIR USES** the specification of which      is attached hereto or   x   was filed on 1/17/97 as Application No. 08/785,532 and was amended on      (if applicable).

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56. I claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

**Prior Foreign Application(s)**

Country	Application No.	Date of Filing	Priority Claimed Under 35 USC 119
			Yes <u>    </u> No <u>    </u>
			Yes <u>    </u> No <u>    </u>

I hereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below:

Application No.	Filing Date

I claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, section 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application No.	Date of Filing	Status
08/731,499	10/16/96	<u>    </u> Patented <u>  x  </u> Pending <u>    </u> Abandoned
08/680,395	7/15/96	<u>    </u> Patented <u>  x  </u> Pending <u>    </u> Abandoned

Full Name of Inventor 1	Last Name Gray	First Name Joe	Middle Name or Initial W.	
Residence & Citizenship	City San Francisco	State/Foreign Country California	Country of Citizenship USA	
Post Office Address	Post Office Address 1921 11th Avenue	City San Francisco	State/Country California	Zip Code 94116
Full Name of Inventor 2	Last Name Collins	First Name Colin	Middle Name or Initial	
Residence & Citizenship	City San Rafael	State/Foreign Country California	Country of Citizenship USA	
Post Office Address	Post Office Address 333 Mountain View Avenue	City San Rafael	State/Country California	Zip Code 94901

(Page 1 of 2)

Full Name of Inventor 3	Last Name <b>Hwang</b>	First Name <b>Soo-in</b>	Middle Name or Initial	
Residence & Citizenship	City <b>Berkeley</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>USA</b>	
Post Office Address	Post Office Address <b>189 Fairlawn Dr.</b>	City <b>Berkeley</b>	State/Country <b>California</b>	Zip Code <b>94708</b>
Full Name of Inventor 4	Last Name <b>Godfrey</b>	First Name <b>Tony</b>	Middle Name or Initial	
Residence & Citizenship	City <b>San Francisco</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>UK</b>	
Post Office Address	Post Office Address <b>699 Teresita Boulevard</b>	City <b>San Francisco</b>	State/Country <b>California</b>	Zip Code <b>94127</b>
Full Name of Inventor 5	Last Name <b>Kowbel</b>	First Name <b>David</b>	Middle Name or Initial	
Residence & Citizenship	City <b>Oakland</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>USA</b>	
Post Office Address	Post Office Address <b>6009 Auburn Avenue</b>	City <b>Oakland</b>	State/Country <b>California</b>	Zip Code <b>94618</b>
Full Name of Inventor 6	Last Name <b>Rommens</b>	First Name <b>Johanna</b>	Middle Name or Initial	
Residence & Citizenship	City <b>Toronto</b>	State/Foreign Country <b>Ontario, Canada</b>	Country of Citizenship <b>Canada</b>	
Post Office Address	Post Office Address <b>717 Bay St. #1501</b>	City <b>Toronto</b>	State/Country <b>Ontario, Canada</b>	Zip Code <b>M5G 2J9</b>

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 1  <b>Joe W. Gray</b>	Signature of Inventor 2  <b>Colin Collins</b>	Signature of Inventor 3  <b>Soo-in Hwang</b>
Date:	Date:	Date:
Signature of Inventor 4  <b>Tony Godfrey</b>	Signature of Inventor 5  <b>David Kowbel</b>	Signature of Inventor 6  <i>Johanna Rommens</i> <b>Johanna Rommens</b>
Date:	Date:	Date: <i>April 2, 1997</i>

Dec-6.mrg 5/95

F:\DATA\PS\PTOD\MERG\72790.1



## DECLARATION

As the below named inventor, I declare that:

My residence, office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **GENES FROM THE 20q13 AMPLICON AND THEIR USES** the specification of which \_\_\_ is attached hereto or x was filed on 1/17/97 as Application No. 08/785,532 and was amended on \_\_\_ (if applicable).

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56. I claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

## Prior Foreign Application(s)

Country	Application No.	Date of Filing	Priority Claimed Under 35 USC 119
			Yes ___ No ___
			Yes ___ No ___

I hereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below:

Application No.	Filing Date


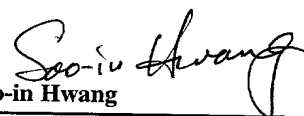
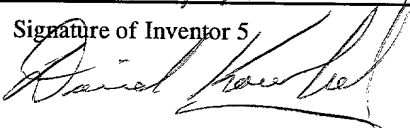
I claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, section 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application No.	Date of Filing	Status
<b>08/731,499</b>	<b>10/16/96</b>	___ Patented <u>x</u> Pending ___ Abandoned
<b>08/680,395</b>	<b>7/15/96</b>	___ Patented <u>x</u> Pending ___ Abandoned

Full Name of Inventor 1	Last Name <b>Gray</b>	First Name <b>Joe</b>	Middle Name or Initial <b>W.</b>	
Residence & Citizenship	City <b>San Francisco</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>USA</b>	
Post Office Address	Post Office Address <b>1921 11th Avenue</b>	City <b>San Francisco</b>	State/Country <b>California</b>	Zip Code <b>94116</b>
Full Name of Inventor 2	Last Name <b>Collins</b>	First Name <b>Colin</b>	Middle Name or Initial <b>Conrad</b>	
Residence & Citizenship	City <b>San Rafael</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>USA</b>	
Post Office Address	Post Office Address <b>333 Mountain View Avenue</b>	City <b>San Rafael</b>	State/Country <b>California</b>	Zip Code <b>94901</b>

Full Name of Inventor 3	Last Name <b>Hwang</b>	First Name <b>Soo-in</b>	Middle Name or Initial	
Residence & Citizenship	City <b>Berkeley</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>USA</b>	
Post Office Address	Post Office Address <b>189 Fairlawn Dr.</b>	City <b>Berkeley</b>	State/Country <b>California</b>	Zip Code <b>94708</b>
Full Name of Inventor 4	Last Name <b>Godfrey</b>	First Name <b>Tony</b>	Middle Name or Initial	
Residence & Citizenship	City <b>San Francisco</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>UK</b>	
Post Office Address	Post Office Address <b>699 Teresita Boulevard</b>	City <b>San Francisco</b>	State/Country <b>California</b>	Zip Code <b>94127</b>
Full Name of Inventor 5	Last Name <b>Kowbel</b>	First Name <b>David</b>	Middle Name or Initial	
Residence & Citizenship	City <b>Oakland</b>	State/Foreign Country <b>California</b>	Country of Citizenship <b>USA</b>	
Post Office Address	Post Office Address <b>6009 Auburn Avenue</b>	City <b>Oakland</b>	State/Country <b>California</b>	Zip Code <b>94618</b>
Full Name of Inventor 6	Last Name <b>Rommens</b>	First Name <b>Johanna</b>	Middle Name or Initial	
Residence & Citizenship	City <b>Toronto</b>	State/Foreign Country <b>Ontario, Canada</b>	Country of Citizenship <b>Canada</b>	
Post Office Address	Post Office Address <b>717 Bay St. #1501</b>	City <b>Toronto</b>	State/Country <b>Ontario, Canada</b>	Zip Code <b>M5G 2J9</b>

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 1  <b>Joe W. Gray</b>	Signature of Inventor 2  <b>Colin Conrad Collins</b>	Signature of Inventor 3  <b>Soo-in Hwang</b>
Date:	Date: <b>5/6/97</b>	Date: <b>5-6-97</b>
Signature of Inventor 4  <b>Tony Godfrey</b>	Signature of Inventor 5  <b>David Kowbel</b>	Signature of Inventor 6  <b>Johanna Rommens</b>
Date:	Date: <b>5/6/97</b>	Date:



I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to:  
Assistant Commissioner for Patents,  
Washington, D.C. 20231,

*May 16, 1997*

TOWNSEND and TOWNSEND and CREW LLP

By *[Signature]*

PATENT

Attorney Docket No. 023070-068920  
UC Case No. 96-185-3

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: )  
)  
Joe W. Gray *et al.* )  
)  
Serial No.: 08/785,532 )  
)  
Filed: 1/17/97 )  
)  
For: GENES FROM THE 20q13 AMPLICON )  
AND THEIR USES )  
\_\_\_\_\_ )

**POWER OF ATTORNEY BY ASSIGNEE  
AND EXCLUSION OF INVENTOR(S) UNDER 37 CFR § 3.71**

Assistant Commissioner for Patents  
Washington, D.C. 20231

Sir:

The undersigned assignee of the entire interest in the above-identified subject application hereby appoints Kenneth A. Weber, Reg. No. 31,677, Kevin L. Bastian, Reg. No. 34,774, M. Henry Heines, Reg. No. 28,219, William M. Smith, Reg. No. 30,223, Ellen Lauver Weber, Reg. No. 32,762, William B. Kezer, Reg. No. 37,369, Tom Hunter, Reg. No. 38,498 and Eugenia Garrett-Wackowski, Reg. No. 37,330, Jonathan A. Quine, Reg. No. P-41,261 as its attorneys to prosecute this application and to transact all business in the Patent Office connected therewith, said appointment to be to the exclusion of the inventors and their attorney(s) in accordance with 37 CFR § 3.71.

An assignment of the entire interest in the above-identified subject application:

☐ was recorded on \_\_\_\_\_ at reel/frame \_\_\_\_/\_\_\_\_.

☐ is submitted herewith for recording.



Joe W. Gray *et al.*  
Serial No.: 08/785,532  
Page 2

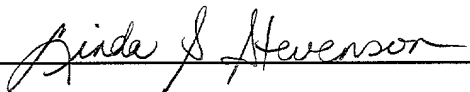
PATENT

Please direct all telephone calls to Jonathan A. Quine at (415) 576-0200 and all correspondence relative to said application to the following address:

Townsend and Townsend and Crew LLP  
Two Embarcadero Center, 8th Floor  
San Francisco, California 94111-3834

Dated: 5/14/97

ASSIGNEE: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA

Signature: 

Typed Name: Linda S. Stevenson

Title: Principal Prosecution Analyst  
Office of Technology Transfer  
300 Lakeside Drive, 22nd Floor  
Oakland, CA 94612-3550

F:\DATA\PS\PTOD\MEG\72907.1

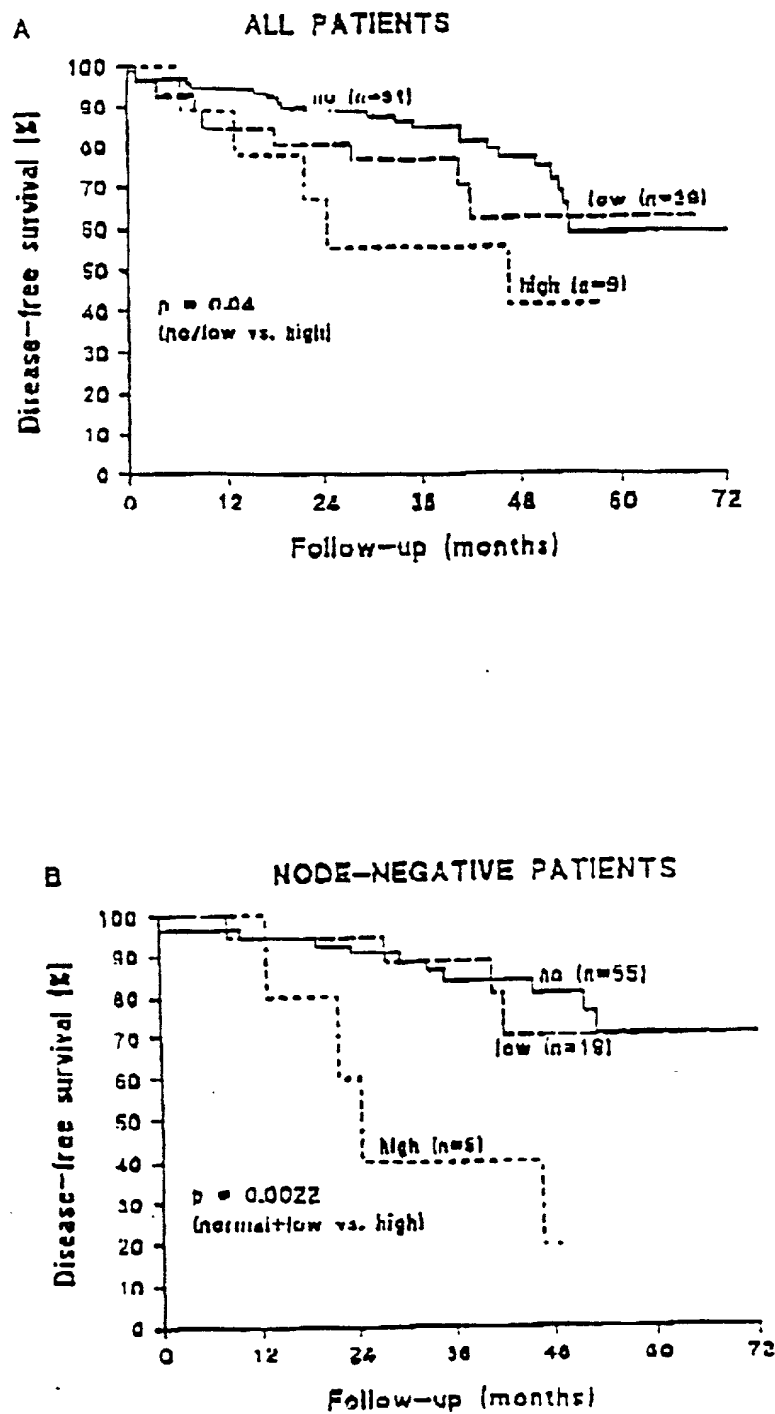


FIGURE 1

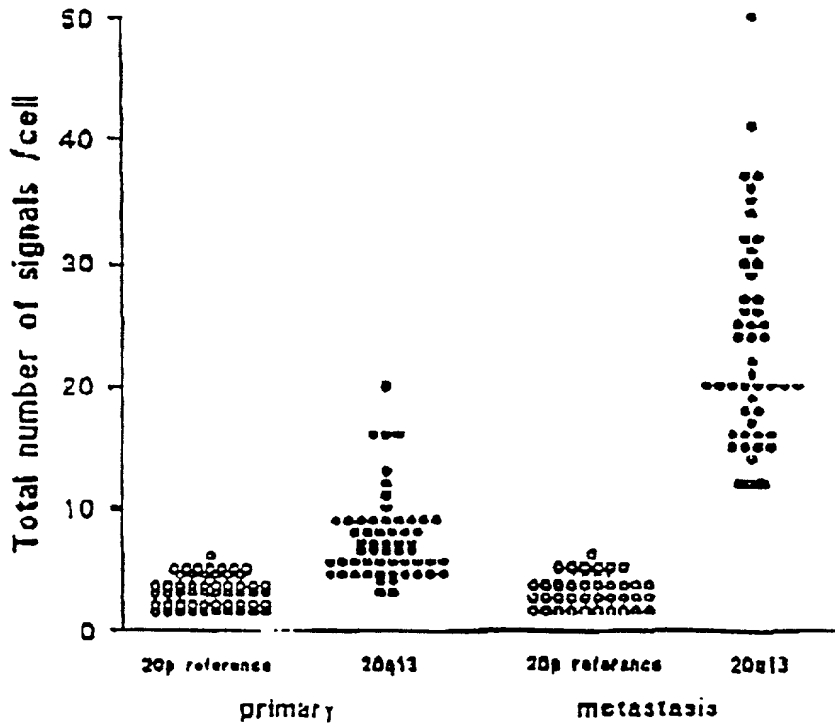


FIGURE 2

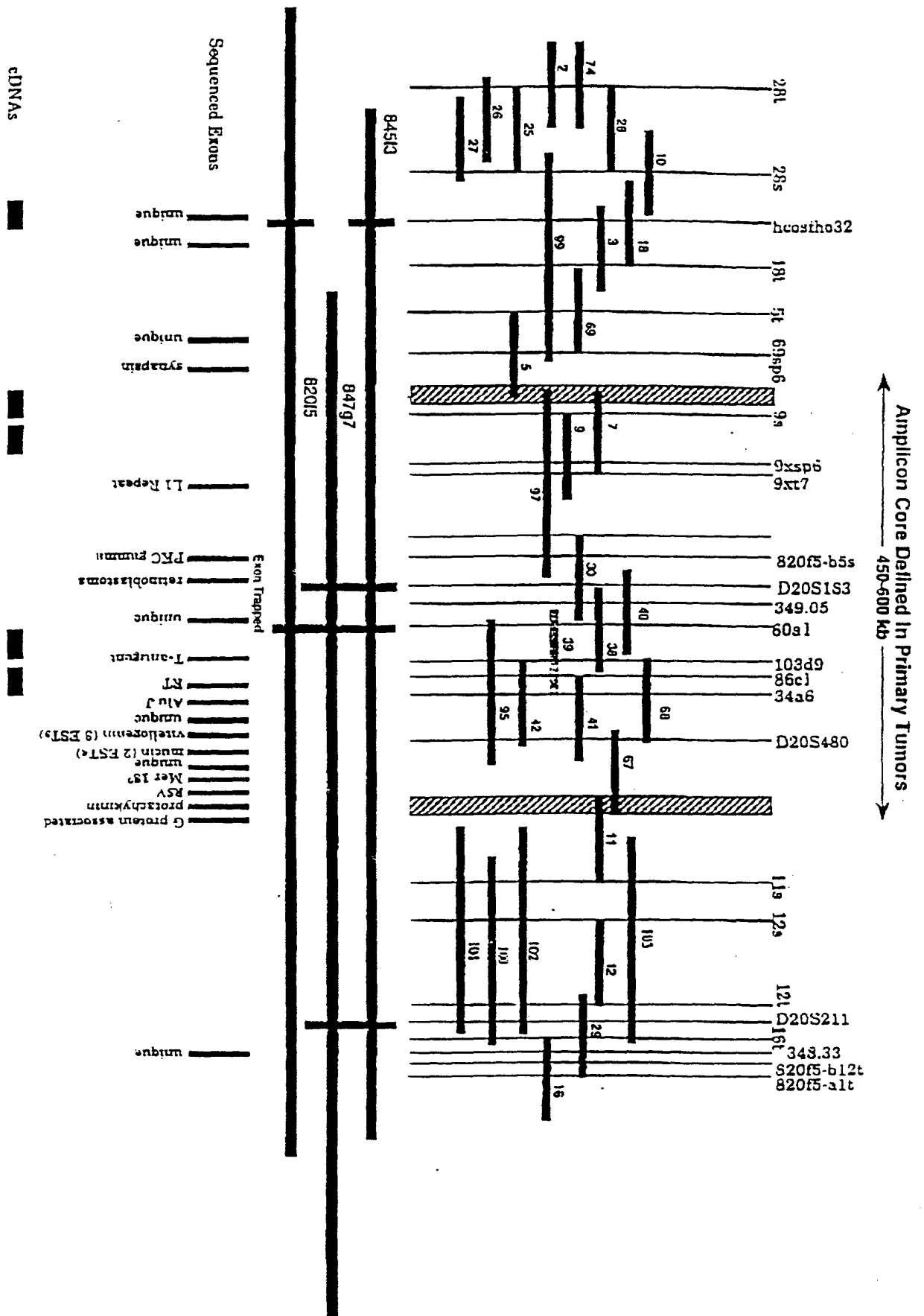
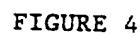


FIGURE 3



Physical map of Q0913.2

RM20C001

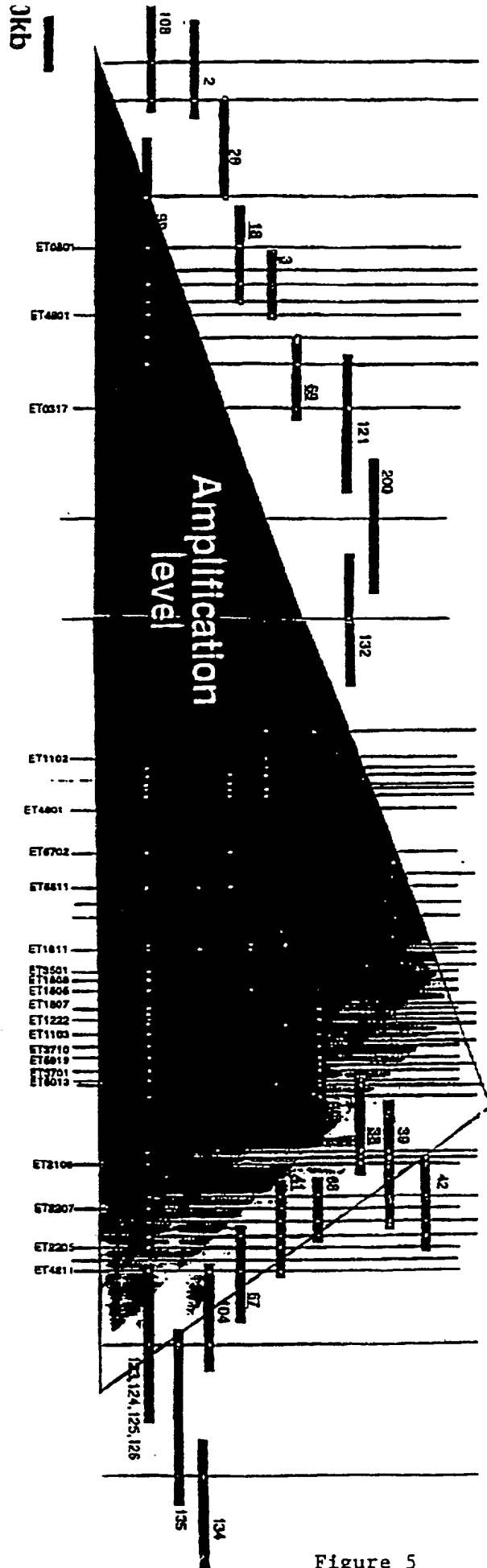


Figure 5

ene ID → 1,2 3 4,5 6,7,8,9,10,11 12,13,14,15

2-90X

08/785532 044757

Filtered query sequence:

```
> query_seq
```

[illegible]

gb|M19533|RATCYCA Rat cyclophilin mRNA, complete cds.  
Length = 743

Minus Strand HSPs:

Score = 418 (115.5 bits), Expect = 1.5e-58, Sum P(5) = 1.5e-58  
Identities = 96/112 (85%), Positives = 96/112 (85%), Strand = Minus /  
Plus

Query: 372 TNCAATATCACCGCAGATGGCGAGCCTTTAGGCCATGTCTCCTTCGAGCTATTTGCAGAC  
313

5-13  
Sbjct: 64 TTCGACATCACGGCTGATGGCGAGCCCTTGGGTCGCGTCTGCTTCGAGCTGTTTGCAGAC  
123

```
Query: 312 AAAGTTCCAAAGACATCAGAAACTTTTCATGGTCTGAGCACTGGAGAGAAAG 261
      |||||
Sbjct: 124 AAAGTTCCAAAGACAGCAGAAACTTTTCGTGCTCTGAGCACTGGGGAGAAAG 175
```

Score = 236 (65.2 bits), Expect = 1.5e-58, Sum P(5) = 1.5e-58  
Identities = 52/58 (89%), Positives = 52/58 (89%), Strand = Minus / Plus

```

Query:      117 TGCTGGACTCACCACAAATGGTTCCAGTTTTTTCATCTGCACTGCCAAAATCGAGTGG 60
            |||
Sbjct:      348 TGCTGGACCAACACAAATGGTTCCAGTTTTTTCATCTGCACTGCCAAGACTGAGTGG
            |||
            405

```

Score = 177 (48.9 bits), Expect = 1.5e-58, Sum P(5) = 1.5e-58  
Identities = 41/48 (85%), Positives = 41/48 (85%), Strand = Minus / Plus

```

Query:      60  GACTGGATGGCAAGCGTGTGATGTTTGGCAAGGTGCAAGAGGGGCATGA 13
             | | | | | | | | | | | | | | | | | | | | | |
Sbjct:     404  GGCTGGATGCCAAGCATGTGCTCTTTGGGAAGGTGAAAAGAGGCATGA 451

```

Score = 154 (42.6 bits), Expect = 1.5e-58, Sum P(S) = 1.5e-58  
Identities = 34/38 (89%), Positives = 34/38 (89%), Strand = Minus / Plus

```

Query:      153 AGAACTTCGTTCTGAAACATGCAGGTCCTGGCATCTTG 116
             |||||
Sbjct:      299 AGAACTTCATCTGAAGCATAACAGGTCCTGGCATCTTG 336

```

Score = 86 (23.8 bits), Expect = 1.5e-58, Sum P(5) = 1.5e-58  
Identities = 22/28 (78%), Positives = 22/28 (78%), Strand = Minus / Plus

```

Query:      256 TCCTGCCTTTGCAGAAATTATTCATTAT 229
              ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct:      193 TCCTCCTTTCACAGAAATTATTCAGGAT 220

```

Figure 6

[illegible]



I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to:  
Assistant Commissioner for Patents,  
Attn: Box Missing Parts, Washington, D.C. 20231,  
on

PATENT

Attorney Docket No. 023070-068920US  
UC Case No. 96-195-3

May 16, 1997  
TOWNSEND and TOWNSEND and CREW LLP

By [Signature]

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	)	
	)	
Gray et al.	)	Examiner: Unassigned
	)	
Serial No.: 08/785,532	)	Art Unit: Unassigned
	)	
Filed: January 17, 1997	)	TRANSMITTAL LETTER - RESPONSE
	)	<u>TO NOTICE OF MISSING PARTS</u>
For: GENES FROM THE 20q13	)	
AMPLICON AND THEIR USES	)	

Attn: Box Missing Parts  
Assistant Commissioner for Patents  
Washington, D.C. 20231

Sir:

Pursuant to the Notice to File Missing Parts of Application - Filing Date Granted dated March 24, 1997, enclosed are the following to be made of record in the above-identified application:

- 1) Executed Declaration (three fully executed Declarations);
- 2) Power of Attorney by Assignee and Exclusion of Inventor(s) Under 37 CFR §3.71;
- 3) Assignment recordation from Joe W. Gray, Colin Conrad Collins, Soo-in Hwang, Tony Godfrey and David Kowbel to The Regents of the University of California; and Assignment recordation from Johanna Rommens to The Hospital for Sick Children;
- 4) Communication under 37 C.F.R. §§1.821-1.825 and Preliminary Amendment;
- 5) The sequence listing in computer readable form;
- 6) Diskette enclosed; and
- 7) Copy of Notice of Missing Parts

06/20/1997 SHOPPER 00000169 DAN:201430  
01 FC:103 594.00 CH  
02 FC:101 770.00 CH  
03 FC:105 130.00 CH  
04 FC:102 160.00 CH

08785532



Gray et al.  
Serial No.: 08/785,532  
Page 2

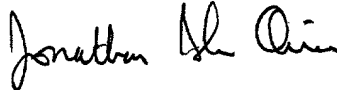
PATENT

Please charge Deposit Account No. 20-1430 for the following fees:

(a) Filing Fee (§ 1.16(a)) (Large Entity)	\$ 770.00
(b) Excess Claims Fees (§ 1.16(b), (c)):	
<u>47</u> - 20 = 27_ x 22.00 =	594.00
<u>5</u> - 3 = <u>2</u> x 80.00 =	160.00
(c) Missing Parts Surcharge (§1.16(e))	130.00
<b>TOTAL FEES TO BE CHARGED</b>	<b>\$1,654.00</b>

The Commissioner is hereby authorized to charge any additional fees associated with this paper or during the pendency of this application, or credit any overpayment to Deposit Account No. 20-1430 for this paper and during the prosecution of this application. This Transmittal Letter is submitted in triplicate.

Respectfully submitted,



Jonathan A. Quine  
Reg. No. P-41,261

TOWNSEND and TOWNSEND and CREW LLP  
Two Embarcadero Center, 8th Floor  
San Francisco, California 94111-3834  
(415) 326-2400  
Fax (415) 326-2422  
JAQ:meg

F:\DATA\PS\PTOD\JAQ\73678.1